

# 数据分析——基于内容与结构的 网络舆情分析报告（2015）

饶 元 冯 妮 宋明爽 员 鹏 睦欣阳 著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

## 内 容 简 介

在 Web2.0 技术与应用不断深化以及移动应用不断产生的新时代,互联网已成为社会民意与民声的重要来源。利用互联网中用户产生的各种数据进行社会舆情分析,已成为当前信息检索与情报获取的一个重要研究方向。人们深刻地认识到,网络信息不仅有研究与应用价值,而且是保证国家或企业安全的重要基础,具有重要的战略意义与社会经济价值。本书对 2014 年天涯论坛网民在经济、政治、民生、生态以及国际与军事等几个关键领域的网络民意,从网络文本数据挖掘的角度进行了初步分析。此外,应用社会网络分析方法直观地展示了一些与内容分析相符的结论。通过这些相关分析,反映出热点事件以及相关民意,期望为相关机构和部门信息决策提供参考。

本书可供社会计算与分析学科、公共政策学科、文本数据统计与挖掘分析的相关专业人员参考使用。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

### 图书在版编目(CIP)数据

数据分析——基于内容与结构的网络舆情分析报告:2015/饶元著. —北京:电子工业出版社,2015.12

ISBN 978-7-121-27808-2

I. ①数… II. ①饶… III. ①统计数据—统计分析—研究报告—中国—2015②互联网络—舆论—研究报告—中国—2015 IV. ①O212.1②G219.2

中国版本图书馆 CIP 数据核字(2015)第 300026 号

策划编辑:甄文全

责任编辑:甄文全

印 刷:

装 订:

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本:720×1 000 1/16 印张:20.75 字数:400 千字

版 次:2016 年 1 月第 1 版

印 次:2016 年 1 月第 1 次印刷

定 价:78.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888。

质量投诉请发邮件至 [zltz@phei.com.cn](mailto:zltz@phei.com.cn),盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

服务热线:(010) 88258888。

# 前 言

网络技术特别是移动互联网络技术与应用的快速发展，推动了大数据时代的来临。一方面，人们越来越多地依赖网络，将其工作与生活的见解、经验、评价与情感以 O2O 的方式即时地共享和发布到网络之中；另一方面，人们也希望将这一些由群体共同创建的信息所组成的大数据资源，通过一些合理的分析方法与共享机制来实现数据信息的价值利用。面对网络中存在的这样一些海量的、结构无序的和以内容为核心的特殊数据，借助社会网络的分析方法与文本挖掘的方法，人们可以更好地从数据之间存在的关系网络结构特征以及内容文本特征中寻找分析问题的新视角。这也是本研究报告在对网络舆情进行分析与研究工作基础之上，希望建构与表达的一种数据分析的整体框架；这也是在 2014 年由科学出版社出版的第一本报告之后，一年来我们对网络舆情研究与分析方法的一次升级与提升。

社会网络分析方法研究是一种相对较为成熟但近年来又十分活跃的领域，该方法原本应用于社会科学领域，通过调查问卷获取采样数据并分析不同的社会结构组成、社会网络中不同组成成员之间的联系、社会结构的演化行为等内容，为人们理解构成社会的人群静态结构特征以及不同类型人群之间的动态行为特征之间的关系，并为管理与决策提供更优的决策依据。

但是，随着网络技术与应用的发展，特别是即时交流工具与 Web2.0 的一些应用，极大地推动了现实物理社会向网络虚拟社会的行为迁移。在这种信息迁移的过程中，人们极大地突破了时间与空间的界限。一方面，使过去以口述自然语言的无记录交流方式，变成了以网络语言的文字或多媒体数据记录的方式进行交流，从而沉淀了大量的文本数据与多媒体数据内容，而且这些数据的规模每年均以指数方式爆炸性增长，从而改变了传统的社会网络分析过程中的数据抽样模式；另一方面，不同的人群沉积的信息中存在大量有价值的信息与知识，如何利用人类历史上最有效的信息共享机制，深度挖掘这些信息中存在的宝贵内容知识资源，为更多的人提供知识与信息服务，已成为新的数据时代下人们面对的一个新机遇与新挑战。因此，社会网络分析也正在向社会计算与社会智能的方向发展与演化。

因此，有人开始对未来网络社会的工作与交流模式进行了预测，并惊呼“淘宝”“微信”等社会化工具的最为令人担心的一件事情就是：随着时间的流逝，

网络记录了一代人的成长、朋友、交流、交易、娱乐等多种类型的个性化数据信息，从而通过这一些数据的分析，很容易形成一个个人或者机构的轮廓模型，通过这个模型可以分析小到普通百姓消费者、大到单位机构甚至国家领导人的行为偏好与情感交流特征，这也是为什么 Google、IBM、Facebook 等国际巨头不惜重金在网络领域与舆情分析领域开展深入研究的根本原因之一。特别是在美国斯诺登曝光的“棱镜门”事件后，使人们更加深刻地认识到，网络信息不仅具有理论研究价值与商业应用价值，同时是保证一个国家或企业安全的重要基础，因此具有重要的战略意义与社会经济价值。

西安交通大学软件学院社会智能与复杂数据处理实验室自 2012 年成立起，便开始从事网络内容与行为的研究，特别希望从文本挖掘与自然语言处理以及社会网络分析方法的角度，探索网络舆情分析方法与社会计算的技术研究策略。三年来，已发布两份社会舆情分析报告，而本报告则是在前面两份报告的基础上，对网络舆情分析方法与指标体系的进一步完善与优化，并对 2014 年 1 月到 2015 年 3 月期间的网络热点事件进行了跟踪与数据分析，期望通过这些相关的分析，找出中国在经济、政治、民生、生态以及国际事务与军事等范畴下的一些关键网络事件，以及人们在事件讨论过程中所折射出来的草根民意，也期望可以为相关机构与决策部门提供一些参考信息。

本报告得到了众多热心人士的支持帮助，特别是西安交通大学软件学院的领导与老师曾经提供了研究环境与设备资源，还有一些老师将关键的技术、业务分析方法以及数据处理的建议毫无保留地进行分享，在此一并道谢。同时，感谢责任编辑为本报告的面世提供了积极的建议与帮助。特别要感谢的是西安交通大学软件学院社会智能与复杂数据处理实验室的所有老师以及研究生们，通过三年来持续的研究与实践，探索了从简单的数据处理到集成舆情分析平台的设计与开发实现，这些工作为后续不断地深入研究与实践提供了大量的技术与人才储备。另外，通过年度报告研究与发布，锻炼和培养一批优秀的青年才俊。这些青年人才及其努力向上的精神是这个实验室最核心的、最宝贵的智力资产与共同精神财富。

本报告由冯妮同学协助进行整体规划，宋明爽、员鹏、睦欣阳参与了报告的指标定义、算法实现以及全文的整理，并负责了相关内容的数据分析，为本实验室的知识服务平台以及金融风险预测与评估平台的建设奠定了相应的技术基础。我们相信，通过实验室所有人员的共同努力，不仅能够为更多的网民、企业以及政府机构提供越来越深入、越来越有价值的文本挖掘与网络舆情分析成果，也可以将相关共性技术应用于不同的研究领域。希望本报告的面世能“抛砖引玉”，建立一个向同行专家学习和交流的新平台，一方面可以不断促进我



们专业分析水平的提升，另一方面可以加强与更多专业机构的协同交流与合作共享。

特别要说明的是，本系列报告的目的在于研究如何利用信息技术与软件工具对网络中的信息进行分析与文本数据的挖掘，也期望利用技术来解决实际中的一些问题。为了保证数据来源的真实性与客观性，在本报告中涉及部分网民的“真实的”网名，在此特别声明，我们仅仅为了表明在研究过程中的客观性，并没有其他的用意。

由于著者的技术水平以及分析能力的限制，本报告虽然较以往的工作有了新的进展，但尚有许多亟待进一步提升与改进的空间，所以希望同行的学者、专家与同仁不吝赐教。希望未来十年，我们可以利用系列报告的方式来不断提升我们的研究与分析水平与能力，以及提高内容与信息挖掘的使用价值，为更多人提供专业的信息与知识服务。

本报告得到了国家科技部“火炬计划”(2012GH571817)、国家社科基金重大项目“基于多学科理解的社会网络分析模型研究——虚拟 Web 网络空间中的社会网络模型与个体行为机制研究”(13ZD177)、2015 年度陕西省科技厅协同创新计划(2015XT-21)“基于网络化感知计算的智慧社区管理关键技术应用示范研究”项目、2015 年西安市科技局产学研协同创新计划(CXY1514(5))“基于内容挖掘的专利地图可视化服务的关键技术研究”项目以及中央高校“科研基金”(08143003)等项目的支持，在此一并表示衷心的感谢。

著 者

西安交通大学软件学院

社会智能与复杂数据处理实验室

2015 年 6 月 25 日



# 目 录

第 1 章 社会网络分析研究概况 .....	1
1.1 基于论坛的网络舆情分析基础 .....	2
1.2 数据特征分析 .....	4
第 2 章 分析方法、模型和数据指标的构建 .....	9
2.1 社会化网络分析方法 .....	9
2.2 主要模型的构建 .....	12
2.3 数据分析指标 .....	17
2.4 主要算法 .....	24
2.5 本章小结 .....	27
第 3 章 领域舆情整体网及数据视图 .....	28
3.1 整体网络特征分析 .....	28
3.2 经济社区网络特征分析 .....	29
3.3 政治社区网络特征分析 .....	36
3.4 民生社区网络特征分析 .....	42
3.5 生态社区网络特征分析 .....	49
3.6 国际和军事社区网络特征分析 .....	56
3.7 本章小结 .....	62
第 4 章 经济领域的相关舆情分析 .....	64
4.1 经济领域热点专题和热点词分析 .....	64
4.2 经济领域意见领袖分析 .....	67
4.3 专题 1: 房市与房价专题 .....	70
4.4 专题 2: 股市与股票 .....	85
4.5 专题 3: 货币与汇率 .....	99
4.6 本章小结 .....	112
第 5 章 政治领域的相关舆情分析 .....	113
5.1 政治领域热点专题热点词分析 .....	113
5.2 政治领域意见领袖分析 .....	117

5.3	专题 1: 国企改革 .....	119
5.4	专题 2: 反腐风暴 .....	133
5.5	专题 3: 依法治国 .....	147
5.6	本章小结 .....	159
第 6 章	民生领域的相关舆情分析 .....	161
6.1	民生领域热点专题和热点词分析 .....	161
6.2	民生领域意见领袖分析 .....	165
6.3	专题 1: 教育与就业 .....	168
6.4	专题 2: 收入分配 .....	183
6.5	专题 3: 社会与医疗保障 .....	199
6.6	本章小结 .....	213
第 7 章	生态领域的相关舆情分析 .....	215
7.1	生态领域热点专题和热点词分析 .....	216
7.2	生态领域意见领袖分析 .....	220
7.3	专题 1: 生态环境 .....	222
7.4	专题 2: 生态资源 .....	236
7.5	专题 3: 技术或措施 .....	249
7.6	本章小结 .....	260
第 8 章	国际与军事领域的相关舆情分析 .....	261
8.1	国际与军事领域热点专题和热点词分析 .....	261
8.2	国际与军事领域意见领袖分析 .....	265
8.3	专题 1: 领土与主权 .....	268
8.4	专题 2: 合作共赢 .....	285
8.5	专题 3: 国际事故 .....	299
8.6	本章小结 .....	312
第 9 章	社会舆情分析工作总结与展望 .....	314
9.1	社会舆情分析总结 .....	314
9.2	下一步工作展望 .....	317
主要参考文献	.....	318
后记	.....	320

## 第1章 社会网络分析研究概况

随着互联网技术的迅速发展与广泛应用, Web 已经成为一个开放的、分布广泛的全球信息服务中心。如何准确、有效地挖掘海量 Web 信息中的价值, 受到越来越多的研究者关注。特别是移动互联网技术的应用, 使得人与人之间、人与群体之间以及群体和群体之间的交互方式发生了巨大的改变, 许多新的社会现象和问题在网络与现实环境的交互过程中呈现出新的动态涌现性、突发及时性、开放共享性、人机之间以及网络与现实之间交互性和数据海量化的特点。这一方面使得基于问卷采样的传统社会网络分析与研究方法, 迫切地需要及时地处理和响应突发事件对社会带来的影响, 以及面对越来越复杂和庞大的社会化网络数据资源来开展新型的问题分析与研究方法; 另一方面, 直接推动了信息科学与社会科学的交叉与融合创新, 如新发展起来的社会信息学 (social informatics)、计算社会学 (computational sociology)、社会人格学 (sociomics)、情报与安全信息学 (intelligence and security informatics) 以及社会计算 (social computing) 等新兴学科, 并反映出多学科交叉与融合创新特点。因此, 利用计算机技术手段与社会网络分析方法, 实现在网络交互环境下海量非结构化数据的深入挖掘与分析, 已成为目前本领域研究的热点。

目前, 社会网络分析中的研究问题主要从三个层次开展, 即整体网络分析、社区网络分析和个体网络分析。整体网络分析侧重于研究网络的整体特性, 如网络的密度、半径和中心度等要素; 社区网络分析强调研究和分析网络中具有相似或相同特征的节点之间所形成的团簇或聚类所形成内部联系紧密而相互之间联系稀疏的网络局部结构; 个体网络分析则侧重于研究个人或组织在其相应的社会网络中所具有的权力、中心地位以及影响力等要素。因此, 通过这三个层次的研究可以有效地进行网络中舆情信息的客观性度量, 从而识别出舆情的整体网络的密度与结构特征、主题事件的动态演化与行为特征、信息的传播特征以及用户参与者之间的关联性结构特征; 在此基础上, 进一步挖掘和发现不同的参与者在整体网络中的度分布特征以及在网络中的权力、影响力和中心地位等特征。通过一些特征的度量与分析, 便于更多的读者利用社会网络分析的方法分析网络中的具体事件和现象, 并解决网络舆情分析过程中的实际问题。

本研究报告的基本定位是: 在去年出版的舆情分析报告的研究工作基础之

上,进一步结合社会网络分析的方法、手段与相关的数据可视化分析工具,来改进和解决舆情分析过程中的一些具体问题,并通过对舆情事件与主题的深入分析,为读者提供了一个更深入的分析结果。

## 1.1 基于论坛的网络舆情分析基础

舆情是民众的社会政治态度与民意的一种综合体现,是社会和谐的晴雨表,它反映了一定时间与社会空间内,围绕特定的社会事件发生、发展和演化过程中,普通民众对社会事件的态度,并通过对事件的态度来映射出对国家管理者的社会和政治态度与倾向。因此,深入了解和分析舆情的产生、发展、演化的变化规律,对于深入了解当下社会发展过程中存在的社会热点、社会问题与社会矛盾等均具有十分重要的作用。

### 1.1.1 关于网络舆情

网络舆情是社会舆情在互联网空间的映射,是社会舆情的直接反映。特别是随着 Web2.0 技术与移动互联网技术与应用的快速发展,人们通过利用网络具有的广泛性、交互性、快捷性和开放性等特点,不仅可以及时了解到新闻时事、社会动态和产品交易,也可以发表个人观点、表达社情民意和公开参政议政。网络已不仅是人们生活与工作的必备工具,也已经渗入到社会的各个领域,成为了各种社会思潮、各种利益诉求和各种意识形态表达其诉求的公共平台。当互联网迅速发展并成为公众社情民意的表达方式并且是较为开放的意愿表达方式时,由于它所具有的自由性与可控性、互动性与即时性、丰富性与多元性、隐匿性与外显性、情绪化与非理性、个性化与群体极化性等特点,所以舆情正面的影响和负面的作用力正在逐渐放大,并对政府和职能管理部门的决策产生越来越大的影响。

近年来的网络突发群体性事件以及网络与现实社会互动条件下所激发的深层社会矛盾,均充分地反映出不同群体对个人利益诉求的追求,以及所引爆的强大网络舆论风险对社会生活产生的巨大影响与直接作用。因此,为了避免不必要的社会风险,解决社会问题与化解社会矛盾,及时识别与发现网络中的突发性事件以及对社会舆情与态度的产生与发展过程进行及时监控和准确判断,已经成为政府正确引导舆情的重要基础,并对政府形象和公信力产生关键的影响。

因此,针对网络舆情中网络、事件、网民、情感、传播互动以及影响力等六个要素,针对公众对自己所关心或与自身利益紧密相关的各种公共事务所持

有的多种情绪、态度和交错意见,系统分析网络中的各种研究对象随着时间演化的特征,对网络事件中的话题进行识别与跟踪、建立网民的情感倾向与影响力图谱以及网民之间的社区关联、并建立网络信息的传播与互动评价模型,从而进一步快速识别网络中的各类事件、动态跟踪热点事件与话题,为相关人员提供网络舆情的实时预警分析与决策,已成为网络舆情分析过程中的核心内容。

### 1.1.2 社会化网络舆情的特点

网络舆情主要是通过开放和虚拟化的互联网进行表达和传播的,它可以通过电子邮件、新闻、BBS、博客、微博和微信等信息渠道快速传播。这些产生于网络上的海量、非结构化的数据,由于具有鲜明的主观性、个性化的、未证实信息数据通过网络直接发布,对数据的分析过程和结果均产生了负面影响和作用。因此,需要通过对不同的数据源进行整合与清洗,从而为社会舆情分析、市场情报分析以及商业智能分析等工作提供重要的数据支撑。但是,由于信息的流动性、及时性越来越高,一些突发的热点事件,很快就会通过网络传播并影响到大量的用户。例如,2015年2月,柴静发布的“穹顶之下”的关于污染与环保的视频,通过微信与微博在极短的时间内获得了上亿次的转发。通过这一个案例可以发现,在新型互联网与移动互联网环境下的“自媒体”时代,网络舆情具有以下特点:

(1) 传播的社区化:网络传播的信息一旦与网民个人兴趣与关注点相同时,通过“选择性接触”极易形成“群内高同质化、群际间异质化倾向”的社区网络;在这种社区网络中,不同的用户在社区中的贡献与重要性不同,导致了网络社区中信息传播方式的不同;另外,社区也会随时时间与关注的焦点问题的不同,而发生动态的演化。

(2) 信息的情绪化与个性化:网络的及时性与方便性,也使得一些网络成为了“自媒体”的渠道。一方面,一些用户遇到个人挫折或是失意,常会通过网络来发泄情绪和对社会的不满;另一方面,用户基于个人对某些事件的片面认识,对特定的社会事件通过网络进行夸大甚至进行造谣,通过庸俗、灰色的情绪化与个性化的言论来引起他人的围观,从而导致信息负面情感的传播以及与信息的实际情况产生偏差。

(3) 信息的偶然性与突发性:网络中的热点舆情事件,往往以实际环境下的偶然突发事件为核心,利用网络的快速传播手段,一个热点话题加上一种情绪化的意见,则可以实现网络虚拟空间与物理空间下的信息互动与快速传播,并在短时间内产生爆炸式的影响力。大量的网络群体性事件均是在这种条件下

爆发和产生的，通过对网络群体的时间演化分析，分析事件的产生、发展、稳定与消亡的过程，从中可以充分反映出广大网民情感、态度、意见和观点的动态演化特征。

因此，把握社会网络环境下的一些关键性特征，对于研究和分析网络环境下的舆情状态，以及一些热点和突发事件进行识别、跟踪、情感判断与预警，具有至关重要的意义，最终实现为舆情管理决策者提供客观的分析依据的目的。

## 1.2 数据特征分析

在数据分析的过程中，数据来源以及数据本身的特征，往往决定了数据分析方法的使用，特别是对于网络舆情的分析过程中，采用传统自然语言处理的方法无法有效地理解有关短文本条件下对内容语义的处理以及大量噪声数据（如水帖和广告帖等）对内容分析结果的影响。由于本报告主要是在网络论坛环境下对论坛中的文本内容进行处理，因此需要对相应的数据特征进行初步的分析与建模。

### 1.2.1 数据源特征分析

Web 数据具有形式多样、发布随意、表达自由等特点，从而使数据分析结果冗余度高、准确度差、数据离散，进而对数据的分析与挖掘过程和结果产生不良影响。因此，如何进行有效地文本数据预处理，消除冗余、去伪存真和关联数据，不仅是确保数据质量的关键所在，也是进一步进行数据分析和挖掘的前提。

作为 Web 数据的一种特殊类型，BBS 网络论坛承载了普通网络民众表达个人观点和意愿的个性化数据内容，其数据具有以下几个关键的特征：

（1）文本数量的规模大。由于目前中文网络论坛数量多，并且中国又具有全世界最多的网民，所以在主流网络论坛上积累的网络文本与信息数量巨大。研究表明，网络热点事件的首发主体中 1/3 左右来自网民，而网民发布信息主要依赖于在线论坛、博客和微博等社交工具，因此对用户参与较多的热门论坛进行研究具有重要的价值和意义。

（2）主题内容复杂，涉及领域广泛。主流的中文论坛几乎涵盖所有领域，不仅包括政治、经济、军事、教育、科技、生态以及国家重要的行业领域，而且还包括社会百态、区域生活等涉及普通民众生活的方方面面。凡是社会生活中被关注问题，都可在论坛上找到相应的讨论区。



(3) 内容表述不规范, 口语化严重。网络论坛的低准入门槛, 使得不同行业、不同文化教育层次、不同社会背景的用户均可以通过论坛发表个性化观点, 内容表述方式不规范, 以生活化与口语化为主, 并且融入许多网络词汇与一些新生的口语化词汇, 如用“酱紫”表示“这样子”, 又如“不明觉厉, 十动然拒”这样的新生词汇, 甚至还存在一些转意的词汇。

(4) 广告与无用信息多。论坛帖子存在大量的广告信息以及“水”帖, 这些信息与主题无关, 噪声数据预处理的质量高低, 直接影响文本数据挖掘和分析的结果。

(5) 信息与事件分布不均衡。论坛中所产生的热点事件具有随机性, 且不同的论坛具有不同的主体特征, 如铁血论坛主要针对军事领域、天涯论坛更多针对社会百态。同时, 同一个论坛中围绕着不同兴趣与倾向的用户提供主题板块的设计, 这对网络热点事件的产生与类型的分布产生了不同影响。

(6) 论坛缺少统一完善的内容分类机制。每个论坛对于文本分类的定义都不相同, 内容相似的帖子也有可能发布在不同的主题板块之中, 如何识别内容并进行过滤和提取, 对于论坛内容和事件的分析也具有重要的影响。

(7) 论坛中用户具有情绪倾向化的特征。网络论坛中的用户大多数属于草根用户, 他们十分关注与个人利益紧密相关的各类事件, 特别容易将对社会的不满, 通过消极和片面的方式体现在网络中, 并对其他的用户产生负面的影响。

因此, 通过对 BBS 论坛这种草根化的在线交流内容的分析和研究, 从中快速地识别出热点事件, 从而可以为不同的用户提供及时的社会热点与动态信息; 同时, 也可以通过对一些突发热点事件的发生、发展、平衡和消失过程的动态跟踪, 清晰地了解到事件的演化过程与行为, 促进相关部门的决策者对当前社会重要事件与舆情状态进行了解与掌握, 从而为舆论引导、危机预防以及管理决策提供必要的数据分析依据。

### 1.2.2 论坛的数据模型

在保证所抽取的数据能够充分地表示论坛中的各类结构与行为特征的前提下, 通过比较分析不同论坛中的数据结构, 本报告从两个角度对相关的数据集进行分析, 并分别以网络中的用户和论坛中发布的帖子内容为核心对象, 研究并建立相应的数据模型。

作为网络中的主体对象, 网络用户通过注册个人的基本信息、在网络中发布的帖子以及对其他的帖子进行回复与浏览等行为, 反映出其个人的一些基本轮廓特征。这些特征不仅包括静态的结构信息, 也包含动态的操作行为以及和网络中的其他用户之间的关联信息。因此, 对用户在论坛中的这些信息的抽取,

可以有效地建立一个用户轮廓模型。另外，结合其他论坛上的用户信息结构与特征，通过重复属性过滤、相近属性的融合等清洗任务，最终形成一个统一的网络用户的属性特征集合。该用户轮廓模型如图 1.1 所示。

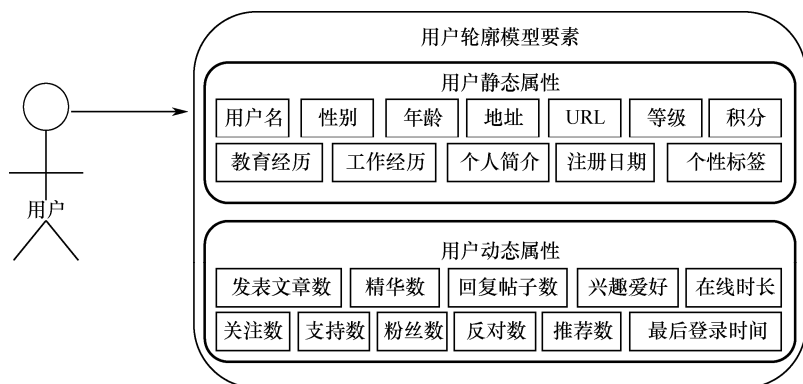


图 1.1 用户轮廓模型

针对论坛中发布的帖子内容对象以及帖子回复内容与关系进行特征描述与分析，将其中的静态属性与动态交互行为进行统一描述和定义，从而形成一个统一的文本内容属性特征模型，为网络文本内容的描述与定义奠定了分析基础。整个网络论坛的内容文本属性模型如图 1.2 所示。通过进一步细分，将帖子的特征要素分成了静态物理属性、静态内容属性、动态内容属性以及动态主体对象属性四类，静态属性反映帖子内容的固有特征，而动态属性反映帖子的交互特性。

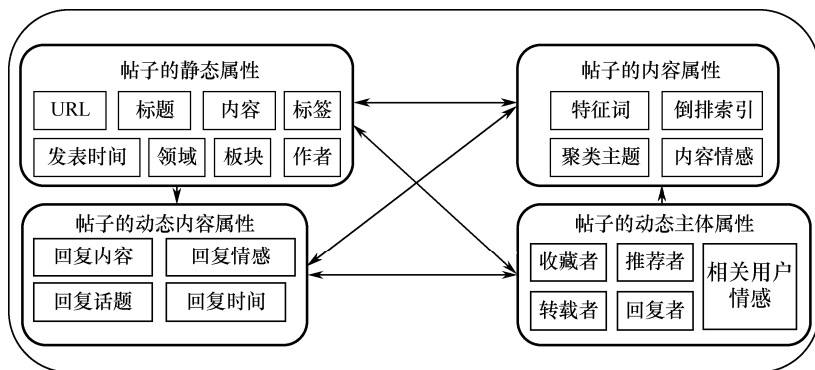


图 1.2 以论坛帖子为主体的文本特征要素模型

由于网络论坛的文本数据是多维属性的集合，所以根据数据集中存在的不同维度和关系，可以构建不同关系类型的社会网络。

### 1.2.3 数据收集和管理

通过对网络论坛文本的属性建模与特征分析,发现这些论坛网站的文本信息几乎都是按照固定的主题板块进行划分的,而且在不同主题板块下均按照发表的时间先后顺序排列。因此,针对这些比较规则的信息,在设计网络爬虫时可以根据网站的顶层 URL 进入网站地图,获取每个板块对应的 URL,将这些板块的 URL 存起来作为爬虫的入口。

网络爬虫包括三个主要模块:第一模块为 URL 收集部分,收集所有要抓取网页的 URL,将其按照顺序存到系统的队列中。第二模块是对网页信息的解析,即针对相应页面上的结构特征与元素特征进行解析。由于网页上的特征信息较多,先过滤掉帖子页面中的广告,并将需要的部分进行内容解析并存储。例如,对于论坛中的帖子,就需要收集帖子的标题、内容、发表时间、所属板块、作者、回复、回复人等信息。第三模块是存储模块,将第二模块获取的数据信息与索引关系存到相应的数据库中。

为了将网络爬虫获取的文本数据有效存储,本研究设计了4个关键的数据库表:数据源表、评论表、用户表以及数据字典。数据源表主要是用于存储抓取的文本数据主体,包括数据 URL、标题、内容、发布时间、来源网站、访问数、回复数、作者、所属板块等关键字段。评论表用于存储每个数据源的评论情况,包括评论内容、评论时间、所属数据源以及评论作者等字段,评论表与数据源表相关联。用户表主要是用来存储发布文章或是发表评论的作者的基本属性信息,如作者名称、URL、所属网站、发表文章数、积分、最后登录时间、注册时间、在线时间、精华数、评论数、个人简介等主要字段。为了保证未来的可扩展性,需要对作者表进行抽象,并增加一个关于博客作者的扩展信息表,通过这张数据表与作者的抽象机制,为数据的抽取与扩展奠定了基础,并形成了如图 1.3 所示的数据存储的底层结构。

另外,为了保证爬虫抓取数据的效率和质量,在设计过程中增加了对重复 URL 以及无法访问 URL 的过滤,并对不同的属性特征进行融合。例如,在有的网站上用户在线时长是按照登录的天数计算,而有的网站上用户在线时长是按照在线的小时计算,为了统一,可以按照 2h 算作一天的规则将其转换成按照在线天数计算。另外,对一些名称不同但是实际意义相同的属性进行合并或转换。例如,有的网站上对于用户观点表示支持的属性为支持数,有的网站同样意义的属性称为“赞”。因此,可以通过对所有类似的同义但不同名的属性进行转换并实现数据的统一,从而保证通过不同页面规则的定义来实现网页内容的自动抓取。

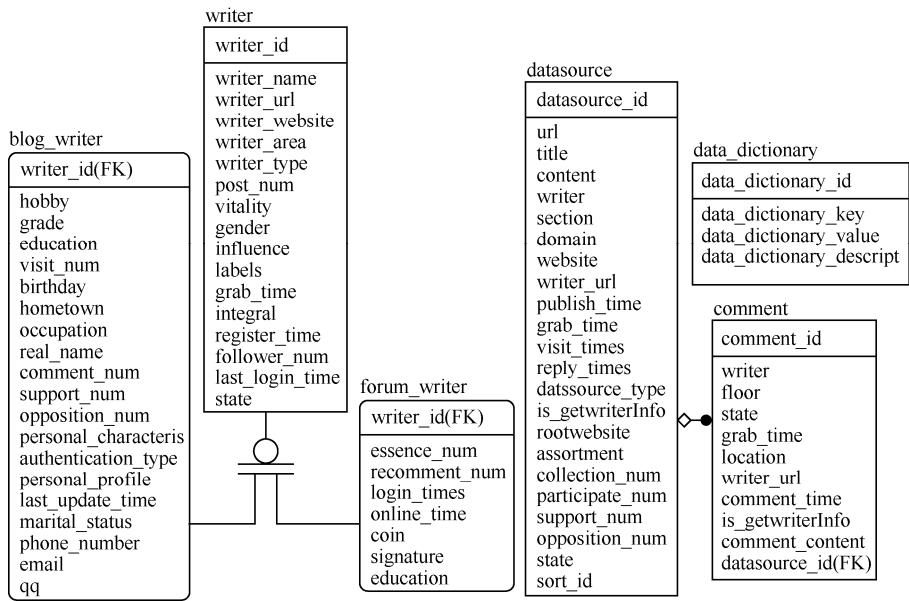


图 1.3 爬虫数据存储的底层结构

本报告在上述模型的基础上，采用基于 **Jsoup** 开源工具开发和实现了一个网络爬虫，并对国内天涯论坛进行数据的抓取和存储。该网站具有大量的用户群体，并且发布的文本内容丰富且具有较大的影响力，有较高的样本示范作用。因此，从该网站上收集到的网络文本数据除对本报告的深入研究提供了基础数据来源之外，还具有较好的内容研究价值与实际应用价值。

## 第2章 分析方法、模型和数据指标的构建

本章在对社会网络分析方法进行分析介绍的基础上，对于网络论坛中的基础分析模型进行初步的构建，并提出本报告所需要分析的指标体系以及相关算法。

### 2.1 社会化网络分析方法

社会化分析软件的广泛应用，使得在物理社会中存在的各类实体关系能在网络虚拟环境下得到映射与体现。通过将社会化活动中的研究对象抽象成一个节点，不同对象之间的关系抽象成一条边，从而将社会关系抽象并形成了一个复杂的网络拓扑结构图。社会领域将社会行动者以及他们之间关系的集合定义为“社会网络”。这种动态变化的、复杂的网络结构，往往可以通过一些网络拓扑关系与特征量，如节点度、最短路径、集聚度、膨胀率、连通性、派系等进行统计和分析，或根据现有的网络模型来研究特征网络的生长与变化，并揭示社会网络的整体结构及其动态演化特征与机制。

由于社会化的内涵包括社交化、社区化和社会化三个不同层次，因此社会化网络也具有三个不同层次：个体网络（ego-network）、社区网络（community network）和整体网络（whole-network）。

个体网络强调以个体对象为中心的社交模式与网络关系结构。例如，腾讯QQ的朋友圈，通过建立个体网络结构，形成一个满足美国社会心理学家Milgram“六度分割理论”的朋友关系“层次”结构，即朋友、朋友的朋友……从而对于社会网络中任何两个互不相识的人，只需要5个中间人就可以建立联系。但这种关系并不是严格的层次结构，不同的节点之间会存在不同的关系，最终形成具有一定“层次性特征”的网络化结构。在社会网络中，大量的事件或话题之间或行动者之间，可能都会存在个体网络结构。例如，网络信息或知识的交换、转发与传播过程所形成的网络结构就是显著的个体网络，该结构是整体网络结构中的一个特例，也反映了不同的分析视角。

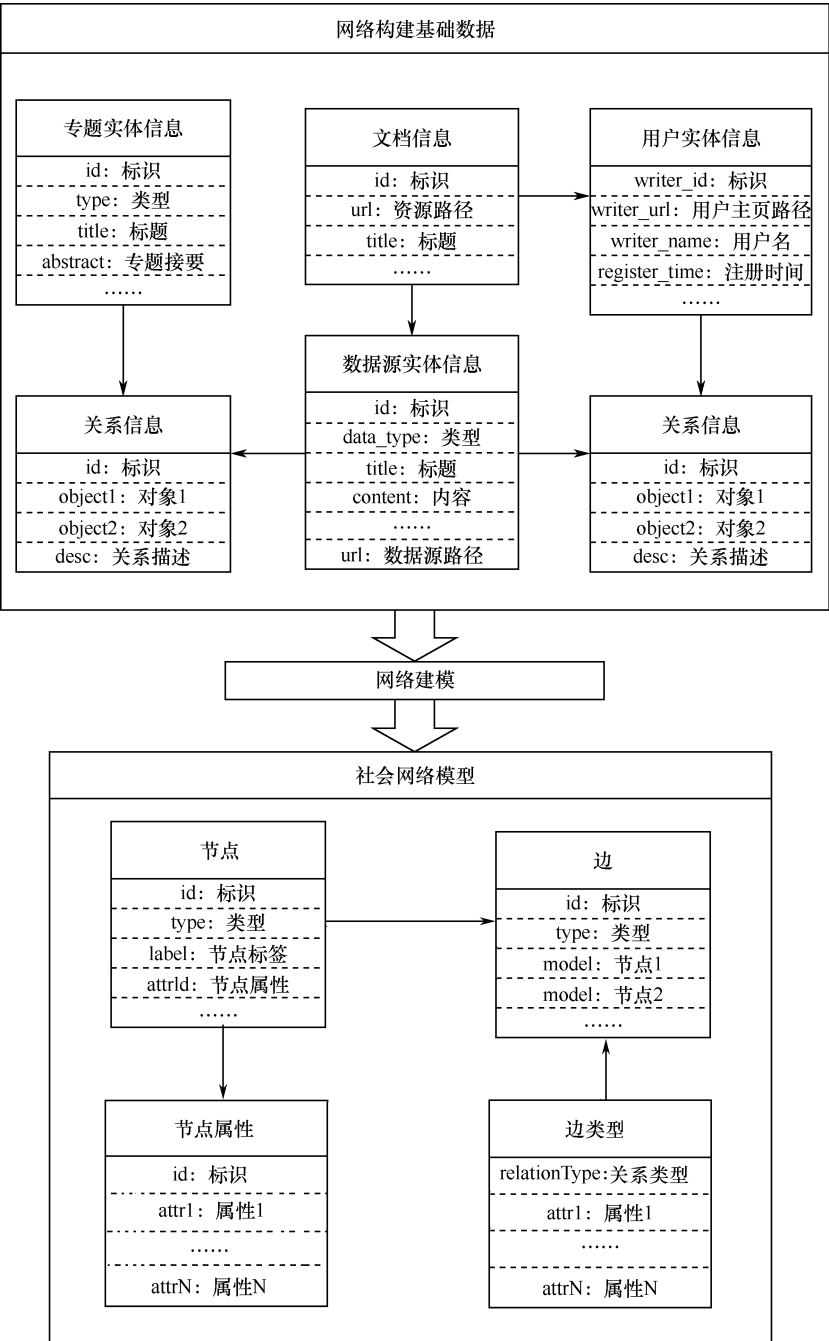
社区网络也是整体网络的一部分，它更强调研究和分析网络中具有相似或相同特征的节点之间所形成的团簇或聚类所形成内部联系紧密而相互之间联系

稀疏的网络局部结构，这种结构也反映了个体行为的局部性特征及其相互间的关联关系，这种关系所存在的局部聚集特征，特别是包含社区的结构特征、社区内群体的演化行为、信息的传播规律以及网络动力学机制等。例如，在目前在线的网络 BBS 社区中，不同的兴趣与爱好的用户，可以通过不同的主题板块来选择具有相关领域的信息来进行互动和交流。腾讯 QQ 上的用户也可以根据不同的个性目标，建立或选择相应的群以实现好友之间的互动与交流，并实现信息的快速分享。而电信公司则可以通过用户通话关系建立动态的社区网络，通过对网络演化的分析，获得客户流失的规律。这些网络中不同特征的社区结构，对整体网络实现了某一个角度或层次的划分。因此，网络社区的特征对于理解整个网络的结构和功能，以及预测和分析整体网络中不同元素间的交互关系起到了关键的作用。

整体网络分析是利用全部数据集对网络的全局结构与整体行为进行总体趋势与宏观决策的把握，主要涉及网络密度和半径、节点的重要性评价、网络社区之间的联系、整体网络的动力学机制，在此基础上对网络的传播与控制、网络博弈产生影响。例如，传播最大化或传播影响最小化问题的研究，就是如何控制信息在网络中的传播范围，以及信息传播过程中所产生的舆情影响力。

在社会网络中，根据不同的研究对象及其特征，可以将网络划分为关于主体对象（行动者所形成的社会化网络）和客体对象（内容实体所形成的社会化网络）。前者关注于人的关系与行为，如电商网络中不同的消费者、论坛中的信息发布者与跟随者等之间的关系与行为；后者关注于具体的某一个实体对象，如书籍、新闻、帖子、博客与微博的内容和评论，根据这些对象的特征以及对象之间存在的关系，来研究内容之间的关系与内容主题之间动态演化行为。

因此，为了获取有效的网络信息，本报告分别以用户为主体的用户关系模型和以文本内容为主体的网络文本模型为基础，通过对社会网络的行动者、文本内容、属性以及关系的描述和定义，经数据处理以及网络信息抽取，从非结构化数据中抽取实体以及实体关系信息，并保存为：文档信息、实体信息（包括数据源实体信息、作者实体信息和专题实体信息）、与实体关系信息（包括数据源—用户关系信息和数据源—专题关系信息）；并利用社会网络的分析方法，将相关的实体、关系与属性特征分别通过网络中的节点、关系边以及不同属性所对应的关系类型，构造不同属性特征下的社会网络模型；并通过整体网络、社区网络以及个体网络三个不同层次，对相关的主体对象进行特征建模与可视化分析。整个分析的过程如图 2.1 所示。



节点

id: 标识

type: 类型

label: 节点标签

attrld: 节点属性

.....

边

id: 标识

type: 类型

model: 节点1

model: 节点2

.....

节点属性

id: 标识

attr1: 属性1

.....

attrN: 属性N

边类型

relationType: 关系类型

attr1: 属性1

.....

attrN: 属性N

图 2.1 内容实体模型转向社会网络模型的实现方法与策略

## 2.2 主要模型的构建

根据拉扎斯菲尔德（Paul Lazarsfeld）提出的二级传播理论，即传播过程至少分为两步：第一级是那些直接接触媒体，对信息的掌握情况相对良好的人；第二级是那些较少接触媒体，主要依靠他人获取信息的人。结合第 1 章中的图 1.1 与图 1.2 提供的基础模型，在对数据源的特点分析基础上，可以将舆情中的用户简单分为两大类：舆情发布者和舆情受众。

### 2.2.1 基于信息发布者的分析模型

在舆情分析中，人作为舆情发布和传播中的主体，是舆情中至关重要的一部分，特别是随着社交网络应用的兴起，生活方式趋向多元化，人们越来越重视个人的真实感受和个性化的思维方式。因此，构建基于网络论坛用户的一个整体网络模型对于分析网民之间的交流关系与倾向，具有重要的意义。

利用社会网络分析方法可以对论坛中主体用户进行抽取与建模，一方面，可以在语义条件下建立不同用户之间存在的共现关系；另一方面，也可以根据不同用户之间存在的属性特征建立一个复杂的关系网络，并通过这些关系网络构建关于用户实体之间的可视化分析模型。这种分析方法如图 2.2 所示。

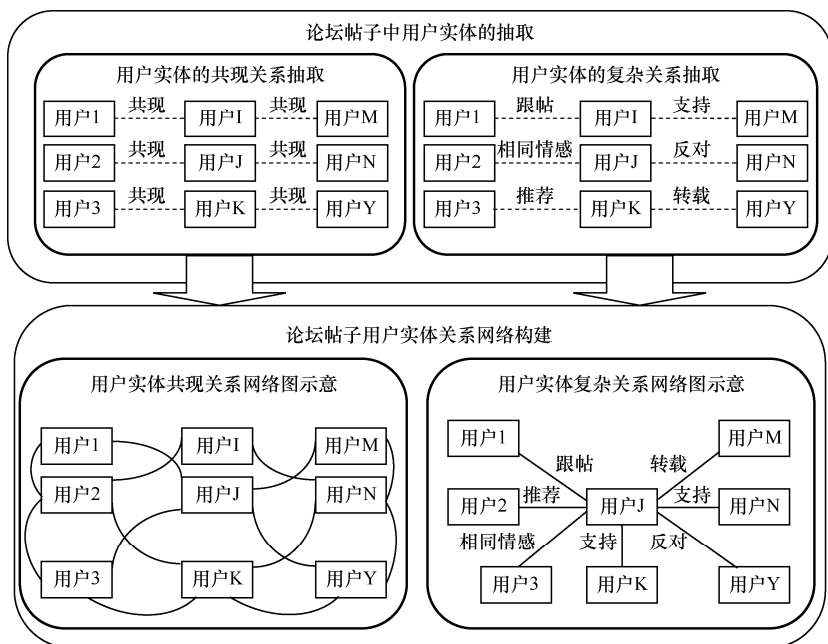


图 2.2 基于社会网络分析的论坛用户关系网络

另外，作为舆情的产生源头，舆情发布者是舆情传播的核心环节，通过对



论坛数据特点的分析,可以从三个方面对舆情发布者进行分析和建模:发布者影响力、发布者活跃度以及发布者的情感倾向。

### 1) 发布者影响力指标

从社会网络分析的视角来看,发布者在整个网络中的中心度大小,直接决定了该发布者节点在网络中的影响能力。为更好地表示这一个指标,作为一个论坛用户,在论坛参与相关的社会化交流的过程中,存在一些显而易见的指标,如所发表帖子的被浏览次数、帖子被回复的次数、帖子的发布时间以及回复的时间等。因此,也可以利用这一些基础的参数来度量论坛中的帖子发布者(舆情发布者),即利用该用户所发布的帖子数目、该用户所发布的帖子被浏览的次数、该用户所发布的帖子所对应的回复数以及该用户的粉丝数四个方面来综合衡量舆情发布者的影响力大小。

### 2) 活跃度指标

影响力指标反映了用户与其他用户之间关联关系的强弱,而活跃度指标则更多强调了用户在网络论坛内部联系过程中的参与程度。在论坛中,可以根据用户注册以来,在某一段时间内参与论坛帖子的交流频次,其中包括了发布帖子数(发帖数)和对其他帖子的回复数(回帖数)。因此,舆情发布者的活跃度可以利用在一定时间段内的发帖数和回帖数这两个指标综合来表示。

### 3) 情感倾向指标

在网络论坛中,一些具有明显的价值与情感倾向的帖子更容易受到网民的关注与传播,进而成为热帖。这些帖子的情感倾向往往反映了该条信息发布者的情感倾向。综合该用户发布的所有帖子进行文本情感的计算与度量,将会从另外一个侧面获得反映用户个人情感的信息,从而判断不同的网民的网络指数感情,区分用户是消极、负面的,还是积极、乐观的,这对于网络论坛的管理和舆情的监测与引导均提供了重要的决策依据。同时,为维护网络论坛的良好交流环境奠定了基础。

## 2.2.2 基于信息参与者的分析模型

舆情受众也是网络论坛中的信息参与者,他们是容易受到网络帖子与其他信息舆情影响的大众网民,他们对舆情发布者所发布的帖子感兴趣,一方面,

进行内容的浏览和阅读；另一方面，也可以对帖子内容发表评论来反映自己对相应事件或问题的看法和建议；甚至可以通过自己对问题的深度分析和评论，以及个人在网络中的影响力来扩大或者改变该信息的舆论导向。因此，结合论坛的一些特征属性，可以从情感指数、参与频度以及受众影响力三个方面来进行度量。

作为对帖子的回复，回帖的内容往往反映了回帖人对原始帖子的支持、反对和中立三种不同的立场倾向。因此，本报告中舆情受众者的情感指数是指该用户参与帖子的回帖中，总的回复数、中性回复数以及情感为正或负的回复数，通过这几个方面的评价，来获得参与人的情感倾向。要特别说明的是，对帖子的回复过程中，需要考虑上下文的语义问题。例如，当原帖的感情倾向为负的时候，回复支持的用户情感在上下文语义的条件下应该也是负向的，而回复反对的用户情感却在上下文语义条件下则为正向的。

针对参与者的影响力而言，可以利用帖子在传播过程中的中心度来进行度量计算，它也反映了网民可能受到舆论情感倾向的影响程度。由于趋向于中立的网民往往容易受到恶意的煽动而改变立场，并转化为对事件负向的情感态度，因此这个指标可以通过中性回复数和参与总数来衡量。因此，一个事件的情感倾向随时间的演化，对于有效了解舆情动态具有重要的分析价值。

### 2.2.3 基于事件的分析模型

利用社会网络分析方法针对非结构化文本数据特征进行抽取建模的过程中，将非结构化的论坛帖子或回复内容作为网络中的节点，反映出不同帖子以及回复之间存在的联系，既可以根据帖子中存在相同或相似的实体来实现文档的并联，也可以根据主题中存在的关键实体，来建立实体与帖子之间的关系，从而形成不同节点与节点之间建立联系的边。整个过程如图 2.3 所示。

网络舆情是在网络的载体下以事件内容为核心，并通过信息的传播与互动，反映了网民的情感、态度、意见、观点以及后续影响力的集合。另外，王飞跃等从量熵、强熵、联熵与情熵四个维度对网络事件进行度量，其中量熵反映了事件在发生与发展进程中，存在的互动或评价的数量信息；强熵反映了事件在发生过程中，单位时间内的互动或评价的数量信息，当单位时间内增加的信息量达到一定程度或阈值时，可能造成的信息快速扩散与舆情事件的突然暴发，因此事件的强熵是舆情分析过程中关键的指标项；联熵反映了事件与其他事件之间存在的关联性强弱，通过关联性分析可能会对主题事件的演化过程提供客

观分析的基础，而情熵则反映了在事件发展演化的过程中网民对事件所反映出来的情感倾向分布与变化趋势。

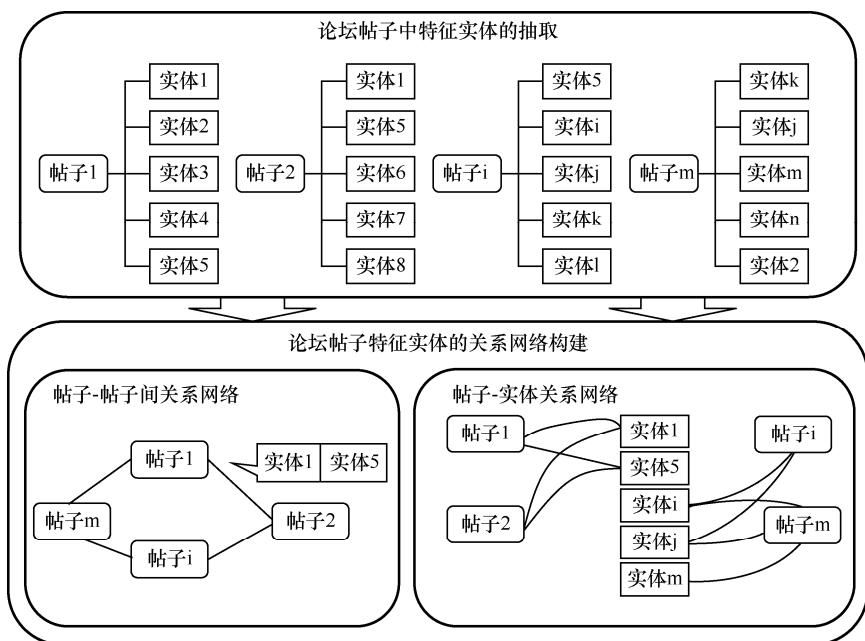


图 2.3 基于社会网络分析方法的事件内容关联分析

舆情热度则是反映了舆情的强度信息，即在单位时间内某一个具体主题相关事件所含帖子的点击浏览量、参与该主题进行发布帖子以及参与回复帖子的总数量；事件情感倾向是通过将事件内所有相关的文本数据进行分词处理后，根据分词的情感极化程度来汇总并统计出来整个事件的情感程度，正向的情感表示积极乐观的，负向的情感表示消极不满的。

最后一个指标是舆情的危害程度，它反映了舆情发展过程中，特别是具有负向的舆情事件对网络社区以及现实社会造成的实际影响。在网络事件引发的群体性公共安全事件多发的情况下，对事件危害程度的评估已经成为一个备受关注新领域。

本报告结合上述的分析指标，从信息主题、舆情热度、事件情感倾向以及危害程度几个方面来建立基于网络舆情的模型。在论坛中，所有事件都针对一个特定的主题域，如执法不公、腐败现象、民生问题、收入差距、国防安全、经济发展、货币政策等主要议题。一些文献将论坛中的主题类别分解为以下一些领域，如表 2.1 所示。

表 2.1 网络舆情信息的主题分类

主题类别	包 括 领 域
生存危机	医疗、信房、就业、教育、社会保障、三农问题、特殊群体
分配差距	城乡差距、区域差距、行业差距、个人收入差距
腐败现象	贪污行贿、挥霍公款、形象工程、唯亲是任、官商勾结、徇私舞弊
公共安全	自然灾害、刑事案件、恐怖威胁、社会治安、民族冲突、黑势力
时事政治	贸易摩擦、政治谣言、党的方针政策
公共治理	决策失误、工商、公检法、税务、城管

由于网民对网络中不同主题领域事件的关注程度不同，并导致了在不同领域下的舆情传播模式也明显差异，明确舆情所对应的主题类别，对于舆情的分析而言，至关重要。

2.2.4 基于舆情传播的分析模型

网络舆情反映了网络事件产生、发展、平衡和消亡的一个过程，在整个过程中，舆情信息的传播则是事件发展和演化的一个基础，由于网络环境下突破了时间与空间的限制，又具有高互动性与去中心化的特征，使得传统的单向传播模式正在向双向或多向的互动传播模式转变，并形成了一种多元复杂网状的形态。

除了利用社会网络中的节点中心度来分析网络中每一个节点在网络生成过程中的作用外，还可以采用舆情传播的几个关键特征和要素来进行度量，其中主要包括了舆情影响力、帖子重现次数、发展过程中的情感拐点和传播强度峰值四个方面。

**舆情影响力**是指舆情的影响与传播程度，它是指网络上特定舆情事件在单位时间内的传播与扩散总量，即包含了事件的点击浏览量、转发量、回复量的总和。**帖子重现次数**反映了同一个主题的事件，在不同网站或在同一网络的不同主题板块内出现的次数，即在不同社区网络中出现的帖子所具有的传播能力；但是，在判断过程中，特别注意到一些水帖和广告内容对分析带来的影响，因此在数据处理前必须先完成数据的预处理任务。事件发展过程中的**情感拐点**是舆情对事件的态度情感倾向转变的临界点，它可以通过对正向回帖和负面回帖的情感差值来进行计算和度量，从而有助于分析大众对事件发展过程中情感变化的拐点与原因，这也反映网络大众的价值取向。**传播强度峰值**是指在单位时间内舆情传统强度达到最大的值，该值代表了舆情事件正在走向稳定或消亡；一般地，当舆情强度达到峰值后，传播的热度就会随之下降。

本报告希望结合社会网络分析方法,对网络论坛中存在的事件进行分析,一方面,研究整个舆情网络中存在的各种结构特征,包括密度、中心势、半径等指标,进而发现网络中的不同对象之间属性特征和行为关系;另一方面,利用在网络论坛中存在的各种特征量,进行指标化的度量和分析,希望可以互为补充,为网络论坛中存在的各类主题事件与话题的分析提供一个新途径、新方法。

## 2.3 数据分析指标

网络舆情监测与预警的分析评价指标体系是以网络媒体技术为手段,展现舆情的分布、来源、传播渠道以及舆情内容的性质和网络大众的情感倾向,并为网络舆情监测和预警机制的建设提供指导思想的评价指标集,旨在实现对非物化社会舆情现象进行可定量或定性的分析能力。

### 2.3.1 指标体系构建

指标是评价某些研究对象时需要确定的评价依据和标准,包括指标名称和数值。指标体系是由一些相互联系、相互补充的指标组成的统一整体,反映了研究对象的综合状况。指标体系以多指标、多层次的方式揭示事物之间的关联性和系统性,将一个复杂的问题分解成多个相联系的部分,通过研究各部分之间的关系,就能透彻认识整体,并准确定位关键制约因素。因此,构建网络舆情指标体系的原则可以归纳如下:

(1) 目标性:对舆情信息而言,常需要通过多个维度来获取相应的信息要素进行分析,以实现一系列的相关目标、满足多方面的实际需求。这就要求在采集、分析舆情信息的过程中协调多个指标。这些指标能够充分体现网络舆情信息的本质,反映网络舆情的典型特征,揭示网络舆情形成模式和变动规律,以便网络舆情工作者对网络舆情进行正确引导和管理。

(2) 科学性:指标体系的设计应当满足定性与定量相结合的原则,即在定性分析的基础上,还要进行量化处理。量化指标需要网络舆情的相关统计数据资料,片面和错误的会导致评估结果出现偏差甚至是误导。对于缺乏统计数据的定性指标,可采用评分法,利用专家意见近似实现其量化。此外,指标体系中同一层次的指标,应该满足可比性的原则,即具有相同的计量范围、计量口径和计量方法,指标取值宜采用相对值,尽可能不采用绝对值。

(3) 系统性: 网络舆情监测与预警工作是一项复杂的系统工程, 涉及诸多方面, 多项指标彼此联系, 能够从多角度、多层次发现网络舆情的特点。网络舆情监测与预警的评价指标体系应当满足以下几点要求: ①相关性: 要运用系统论的相关性原理不断分析, 然后设计评价指标体系。②层次性: 指标体系要形成阶层性的功能群, 层次之间要相互适应并具有一致性, 要具有与其相适应的导向作用, 即每项上层指标都要有相应的下层指标与其相适应。③整体性: 不仅要注意指标体系整体的内在联系, 而且要注意整体的功能和目标。④综合性: 指标体系的设计不仅要有反映事故状况的指标, 更重要的是要有反映隐患的指标, 事前与事后综合, 不同时期(历史、现状、将来)综合才能更为客观和全面。

(4) 定量性: 指标的描述要求概念明确、简单明了, 还要考虑现行科技水平和可行的网络舆情信息的收集方法与渠道, 以实现有效的采集数据; 尽量选取易量化的指标, 减少主观指标的数量。

(5) 可操作性: 构建指标体系是为了实际操作应用, 应该尽可能做到简单、精练, 方便检测、采集及评价。简单、精练的评价指标体系, 能够在不影响评价结果的条件下缩短并简化计量、处理、评定等测评工作, 提高评价的工作效率。

### 2.3.2 分析指标

依据指标建立的几个方面的特征, 将对论坛中的内容分析设定一个科学的指标分析结构体系, 并利用该指标体系建立一个更易于定性理解或定量分析的网络舆情信息评测机制, 对研究内容的分析与挖掘提供一个客观依据。本书构建的舆情分析指标体系包含了基于内容和基于结构的两方面的指标。本书的舆情分析指标体系如图 2.4 所示。

#### 1) 基于内容的指标

从网络舆情的传播过程来看, 可以从以下几个方面来描述网络舆情: 舆情传播者、舆情要素、舆情受众、传播属性和舆情爆发区域。因此, 基于内容的指标体系包含舆情发布者、舆情要素、舆情受众、舆情传播、区域和谐度五个部分。

##### 1. 舆情发布者指标

舆情发布者指标主要用来度量发布者引起网络舆情分析者受关注的程度, 是监测具有号召力与舆论影响力网民的特征性指标, 如评定和度量网络中的“论坛领袖”。它包括以下关键指标:

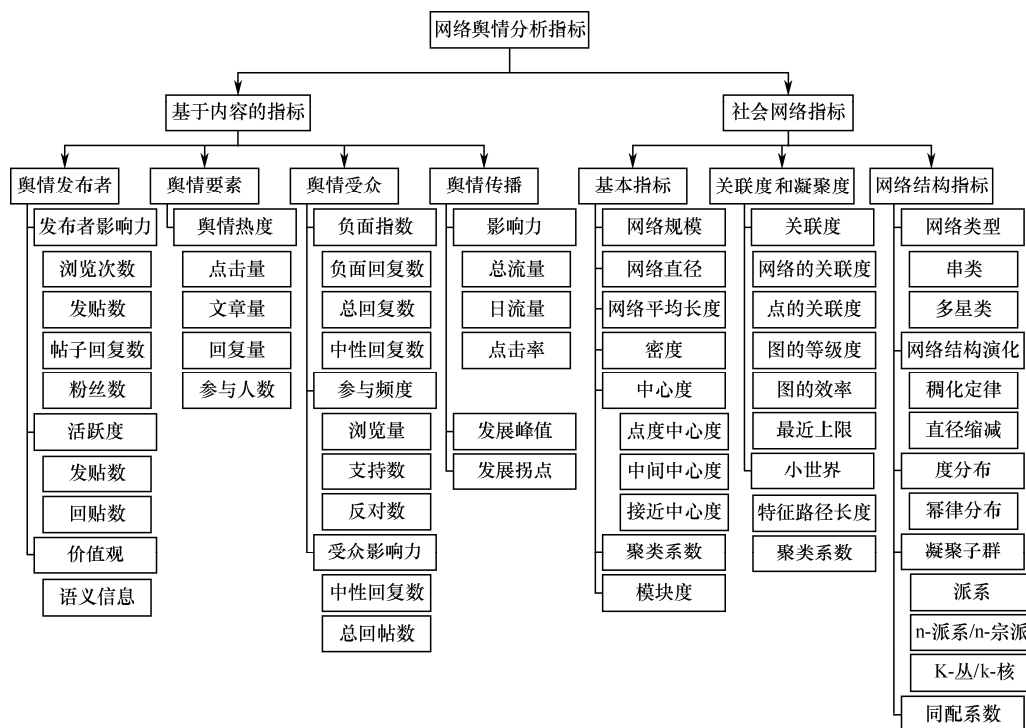


图 2.4 网络舆情分析指标体系

**发布者影响力：**用来度量发布者在社区的作用与影响力的大小。该指标通过浏览次数、发贴数、帖子回复数和粉丝数四个网络参量来进行度量。

$$N = g_2 N_2 + g_3 N_3 + g_4 N_4 \quad (2-1)$$

$$AI = 2 * \text{sigmoid}((g_1 * N_1 / \text{Max}(N_1) + g_5 * N / \text{Max}(N)) * 6.0) - 1 \quad (2-2)$$

式中， $AI$  发布者影响力， $N_1$  为总文章被浏览次数， $N_2$  为发布文章数， $N_3$  文章被回复数， $N_4$  为粉丝数，权重分别为  $g_1$ 、 $g_2$ 、 $g_3$ 、 $g_4$ ； $N$  为中间计算结果， $g_5$  为其所占权重。在计算作者影响力中用到了  $\text{sigmoid}$  函数，即

$$\text{sigmoid}(x) = 1 / (1 + e^{-x}) \quad (2-3)$$

由于当  $x=6$  的时候该函数已经非常接近 1，而且随着  $x$  继续增大该值变化特别小，所以通过归一把  $x$  控制在  $[0,6]$ ，因为  $\text{sigmoid}$  函数的值域为  $[0.5,1)$ ， $2 * \text{sigmoid}(x) - 1$  的值域为  $[0,1)$ ，故最后的函数定义如式 (2-3) 所示。

**发布者活跃度：**用来度量发布者在此社区的活跃程度，可以通过该用户在网络中的发贴数和参与的回帖数来衡量，并可以利用下面的公式计算：

$$N = g_1 * N_1 + g_2 * N_2 \quad (2-4)$$

$$AC = (N / \text{Max}(N)) \quad (2-5)$$

式中,  $AC$  为发布者活跃度,  $N_1$  为发表的文章数,  $N_2$  为参与回复数, 权重分别为  $g_1$  和  $g_2$ 。

**价值观:** 是指在网络世界中, 舆情发布者发布的信息中体现出来的对事物的评价和态度取向。

## 2. 舆情要素指标

舆情要素指标主要强调观测舆情发展的深化过程与整个主题的全生命周期, 主要包括三种: 舆情热度、舆情的主题分类和舆情的危害程度, 其中舆情热度指标是衡量舆情的最关键指标。

**舆情热度:** 主要用来度量舆情受网络用户关注的程度, 并可以通过网络参数——点击量、文章量、回复量和参与人数来进行组合计算, 其中公式如下:

$$N = g_1 * N_1 + g_2 * N_2 + g_3 * N_3 \quad (2-6)$$

$$CH = 0.5 * (g_4 * N_4 / \text{Max}(N_4) + g_5 * AN / \text{Max}(AN)) \quad (2-7)$$

$$t = 6.0 * t_{ef} / t_{\max Ef}, \text{ 当 } t > 0 \text{ 时,}$$

$$CH = (g_4 * N_4 / \text{Max}(N_4) + g_5 * AN / \text{Max}(AN)) / t_{ef} * (1 + e^{-t}) / 2.0 \quad (2-8)$$

式 (2-6) ~ (2-8) 中,  $CH$  为舆情热度,  $N_1$  为文章数量,  $N_2$  为回复数量,  $N_3$  为参与人数, 相应权值为  $g_1$ 、 $g_2$ 、 $g_3$ ;  $N_4$  为点击量,  $N$  为中间计算结果,  $AN$  为有效天数的中间计算结果总和,  $g_4$  为点击量所占权重,  $g_5$  为中间计算结果所占权重。

若某一天的  $N$  大于所设定阈值, 则改天为有效天数,  $t_{ef}$  为有效天数总和, 进行归一化处理后, 得到最终结果。

**信息主题的危害程度:** 通过文本内容的分析, 按照一定的标准, 获得舆情事件的危害程度的一种分析方法。

## 3. 舆情受众指标

舆情受众指标主要针对所有被舆情影响到的网络用户 (网民), 它主要包含以下三个子指标: 舆情负面指数、参与度和受众影响力。

**舆情负面指数:** 描述的是舆情对网民思想冲击而产生的负面情绪的程度。

$$NI = (g_1 * N_1) / N \quad (2-9)$$

式中,  $NI$  为舆情负面指数,  $g_1$  为相应权重,  $N_1$  为负面回复数量,  $N$  为回复总数量。

**参与度:** 指网民参与舆情讨论的程度。

$$PF = RN / N \quad (2-10)$$

式中,  $PF$  为参与度,  $RN$  为回复数量,  $N$  为点击数量。

**受众影响力:** 用来度量网民可能受到负面舆论影响的程度。因为趋向于中



立的网民往往容易受到恶意的煽动，从而转化为对事件的负面态度。

$$PI = NN / N * \lg(10 * N / AN) \quad (2-11)$$

式中， $PI$ 为受众影响力， $NN$ 为中性回复数量， $N$ 为回复总数量， $AN$ 为所有事件的平均评论数量。 $\lg(10 * N / AN)$ 避免总回复量较小而受众影响力较大的情况，同时也避免了总回复量较大而造成其影响力特别大的情况。

#### 4. 舆情传播指标

舆情传播类指标主要目的在于确定舆情在网络中的传播渠道和传播影响的环节，并着重可以分析网络舆情所处的传播和扩散阶段。它包括以下四个子指标：

**舆情影响力：**主要是指舆情传播过程中对网民的影响程度，通过某一专题下所有帖子的日流量和总流量随时间变化的趋势图来衡量。日流量即所有帖子的日回复量之和，总流量即所有帖子的总回复量之和。

**发展峰值：**舆情讨论最激烈的点，即单位时间内回复量最高点（发展峰值），监控该点可以得知导致舆情变化的原因与趋势。

**发展拐点：**即情感倾向发生变化的点，通过正面回帖数与负面回帖数的差值来进行计算。监控该指标可以有助于分析引发网络大众情感发生变化的拐点与原因。

**帖子重现次数：**报道次数指的是该舆情信息在传播过程中被报道或被发布的次数。

#### 5. 区域和谐度指标

区域和谐度指标是从区域差异的视角来考察潜在的社会矛盾，该地区社会矛盾越突出，越容易刺激网民的情绪，它包括贫富差距、信息沟通、社会保障和宗教信仰。信息沟通指的是社会信息传播交流系统，是人与人、人与社会了解彼此的观点、态度、决策的重要渠道。宗教信仰考察的是邪教、民族矛盾与宗教信仰。宗教问题形成后，往往带有一定的倾向性，加大了各级政府行政管理的难度，危害到当地的社会稳定与经济发展。

### 2) 社会网络分析指标

本书中，主要使用社会网络分析法对舆情的传播过程中形成的网络结构进行分析，因此，基于结构的舆情指标也就是社会网络分析指标。从指标表征的含义来看，该指标体系可以分为社会网络基本指标、关联度和凝聚度和网络结构指标三个部分。

#### 1. 基本指标

基本指标是对于整个网络的一套通用指标计算标准。任意一个网络都是由

节点和边两部分组成。网络中节点间的距离指的是从一个节点出发达到另一个节点要经历的边的最小数目。

**网络规模：**网络的大小。通常由构成该网络的节点个数来表示。

**网络直径：**任意两个节点间距离的最大值。

**网络平均长度：**网络中所有节点对之间的平均最短距离。

**密度：**是一个图各点之间联络的紧密程度。固定规模点之间的连线越多，则说明该图的密度越大。具体来说，是指在网络中实际存在的线与可能存在的所有的线的比值。用来测量社会网络中行动者之间的联系程度，值介于[0,1]。

**中心性：**指的是位于中性或具有中心的性质。中心性主要可以分为三个方面来衡量：点度中心性、中间中心性和接近中心性。每种中心性都有对应的中心度和中心势两个指标。中心度是描述图中某一节点居于某种中心的衡量指标。中心势是描述整个图的紧密程度或一致性，也就是一个图的中心度。

**中心度：**描述图中任意一点在网络中占据的核心性。中心度主要可以分为三个方面来衡量：点度中心度、中间中心度和接近中心度。点度中心度又包括绝对中心度和相对中心度，绝对中心度就是指该点的度数，相对中心是绝对中心度与图中点的最大可能的度数之比。中间中心度是用来测量社会行动者对资源控制的程度。接近中心度表示一种不受他人控制的程度。

**聚类系数：**是表示一个图形中节点聚集程度的系数。证据显示，在现实中的网络中，尤其是在特定的网络中，由于相对高密度连接点的关系，节点总是趋向于建立一组严密的组织关系。

聚类系数表示的是与节点  $i$  相连的  $k$  个点， $k$  个点之间的实际存在边与总的  $k$  个节点的所有可能的边的比值。

$$CC_v = \frac{n}{C_k^2} = \frac{2n}{k(k-1)} \quad (2-12)$$

**模块度：**评估网络某种特定划分的质量的度量。 $Q$  就是模块度的衡量标准。 $Q$  值越大，表示该网络具有很强的社区结构；在实际网络中，该值通常位于 [0.3,0.7]。

单个社区模块度的定义：

$$Q = \frac{1}{2m} \sum_s (e_{ss} - a_s^2) \quad (2-13)$$

该指标的基本假设是：随机网络没有社区结构；如果原始网络的一个子网络内部的边数比其在相应的随机网络中的期望边数高，则认为该子网络对应原始网络的一个社区；“期望”是指在原始网络的所有可能随机化网络上进行平均。

模块度的提出为研究人员提供了一个用户选择网络划分的目标函数。

## 2. 关联度和凝集度

关联度是指集体成员之间相互联系的程度。凝集度指的是一个集体的全部成员通过社会关系联系在一起的程度。关联度的衡量可以分为以下几个指标。

**网络的关联度：**通过“可达性”来测量关联程度。关联性的测量公式为

$$C = 1 - \left[ \frac{V}{N(N-1)/2} \right] \quad (2-14)$$

式中， $V$  是该网络中不可达的点的数目， $N$  是网络的规模。关联度的计算运用可达矩阵。

**网络中点的关联度：**两点之间的路径越多，关联度越好。

**图的等级度：**是指人们相互之间在多大程度上非对称地可达。图的等级度计算公式为

$$GH = 1 - \frac{V}{\max(V)} \quad (2-15)$$

式中， $V$  是网络中对称地可达的点对数。 $\max(V)$  是  $i$  达  $j$  或者  $j$  可达  $i$  的点对数。

**图的效率：**是指在已知图中所包含的成分数确定的情况下，图在多大程度上存在多余的线。图的效率的计算公式为

$$GE = 1 - \frac{V}{\max(V)} \quad (2-16)$$

式中， $V$  是多余线条数， $\max(V)$  是最大可能多余线的条数。

**最近上限：**关注的是树形图中存在多少个“根”。对于任意一个行动者来说， $LUB$  就是能够达到二者的最接近的人。 $LUB$  的计算公式为

$$LUB = 1 - \frac{V}{\max(V)} \quad (2-17)$$

式中， $V$  等于网络中不存在  $LUB$  的点对数，这个值的计算要根据全部成分进行汇总；分母为

$$\text{Max}(Vn) = \frac{(N_n - 1)(N_n - 2)}{2} \quad (2-18)$$

**小世界：**网络图中大部分的节点不与彼此邻接，但大部分节点可以从任一其他点经少数几步就可到达。若将一个小世界网络中的点代表一个人，而连接线代表人与人认识，则这小世界网络可以反映陌生人由彼此共同认识的人而连接的小世界现象。

如果对小世界现象进行形式化的表征，对于互联网来说，可以利用如下两

个统计量来刻画其性质。

(1) 特征途径长度  $L$ : 连接两个点之间最短途径的平均长度。

(2) 聚类系数  $C$ : 聚类系数是一种关于局部网络结构的指标, 该指标有两种界定方式: ①平均局部密度。如果用  $C_v$  表示点  $v$  的聚类系数, 那么  $C_v = v$  的点的个体网密度, 整个网络的聚类系数等于各点密度系数的均值, 即  $C = C_v/n$ 。

②传递性比率: 封闭三方组的数量与三方组总量之比。

### 3. 网络结构指标

**网络类型:** 分为串类和多星类。不同结构类型的网络指标经常有很大区别, 信息的传播方式也不同。分析网络类型有利于发现网络传播模式。

**网络结构演化:** 在社会网络结构的演化中有两条重要的特性——稠化定律和直径缩减定律。稠化定律指的是网络中边的数目相对于节点的数目以超线性速度增长; 直径缩减定律指的是网络的有效直径随着网络的增长是减小的。

**度分布:** 节点的度数服从的概率分布。最典型的度分布是幂律分布。

**同配系数:** 描述网络中的节点和与其度相同的节点连接的倾向性; 若该指数大于 0, 网络是同配的, 表示节点倾向于和与其度相同的节点连接; 若该指数小于 0, 网络是异配的, 表示节点倾向于和与其度相异的节点连接。

**凝聚子群:** 群体是在既定目标和规范的约束之下, 彼此互动、协同活动的一群社会行动者。凝聚子群是满足如下条件的一个行动者子集合, 即在此集合中的行动者之间具有相对较强、直接、紧密、经常或积极的关系。对凝聚子群的分析主要有四个方面: ①关系的互惠性。②子群成员之间的接近性或可达性。③子群内部成员之间关系的频次(点的度数)。④基于“子群内外关系”的凝聚子群。这四个不同角度的研究分别对应了不同的凝聚子群概念。建立在互惠性基础上的凝聚子群主要是派系; 建立在可达性基础上的凝聚子群是  $n$ -派系和  $n$ -宗派; 基于度数的凝聚子群是  $k$ -丛和  $k$ -核; 基于“子群内外关系”的凝聚子群有成分、块、LS 集合等。

## 2.4 主要算法

大数据下的文本挖掘会牵涉很多的算法, 通过这些算法, 可以将一些结构化或半结构化的文本数据抽取我们想要的一些数据, 以达到分类、聚类、摘要或关键字的抽取、情感计算等问题的目的。

### 2.4.1 数据分类

通过数据分类可以把一个未知领域的数据进行领域分类, 然后对某个领域

可以进行具体的舆情分析，其主要步骤如下：

- (1) 对已分词的文本进行词频统计，取出前 2000 个关键词。
- (2) 算出每个词的不同领域的先验概率。
- (3) 根据贝叶斯分类方法对文本进行分类。

## 2.4.2 专题聚类

通过专题聚类可以把一个领域下的所有帖子聚类成多个专题，而对于回复数较多的专题即为热门专题，也就是网友比较关注的专题，分析这些专题可以准确地把握某一个领域下的舆情分类信息。针对文本的专题聚类算法有很多，本书中的专题聚类算法采用了结合层次聚类的密度聚类方式，其主要步骤如下：

- (1) 对在领域下帖子的标题和内容进行分词并加权，建立文档的倒排索引。
- (2) 采用自顶向下的分裂聚类算法，因此需要给出第一层 DBSCAN 聚类的  $\varepsilon$  邻的最大值。使用 DBSCAN 聚类算法进行第一层聚类。
- (3) 修改  $\varepsilon$  邻域半径的取值，使用 DBSCAN 聚类算法继续下一层聚类。
- (4) 如果  $\varepsilon$  邻域半径小于最小半径，层次聚类结束，否则返回第三步。
- (5) 取专题下频率最高的 100 个非停用词作为该专题的关键词，并过滤形成专题的特征向量。

通过以上步骤，将得到不同领域半径下的帖子聚类结果，每个分类下的第一层聚类得到该领域中几个领域主题，通过结合其他指标的筛选将得到领域中的热点主题。后面每层的聚类结果都是上一层主题包含的子主题。

## 2.4.3 专题标题抽取

专题的标题一般表示专题的关键特征所在，即需要机器自动地找到某一个最能够反映出专题特征帖子的标题，并将其作为该专题的候选标题。因此，通过针对专题内所有帖子的标题特征抽取，就可以了解在该专题中将会涉及哪些关键特征信息。专题内相关标题的特征抽取算法步骤如下：

- (1) 根据第一步得出的事件特征向量进行处理，得到高频词的  $TF*IDF$  值。
- (2) 取出一个事件下的所有的帖子，并且取出其标题。将每个标题与对应的高频词进行比对，得到出现高频词最多的一条帖子的标题。
- (3) 将每个标题与对应的高频词进行比对，如果句子中含有高频词，则给句子打分加 1，依次计算并记录每个句子的打分情况。
- (4) 最后，将得到的所有句子的打分进行排序，得到打分最高的帖子标题，将其作为事件的标题。

### 2.4.4 专题内容摘要的自动抽取

专题内容简介的自动抽取是指对特定专题的内容进行摘要抽取，即通过对专题内容中常出现的一些信息进行自动化抽取，并自动形成专题的摘要信息。通过查看专题内容摘要便可以知道该专题主要涉及哪些关键事项，从而有助于更好地去查看专题下的所有帖子和回复的详细内容。一般地，专题内容摘要的抽取算法的主要步骤如下：

(1) 抽取对应事件下的所有的帖子，并抽取其内容。

(2) 将某个事件下对应的帖子的所有的内容集合与聚类所得的高频词进行比较，得到含高频词最多的 5 个帖子内容的集合。

(3) 运用 PageRank 的思想，运用 BM25 算法，提取出五个句子中摘要，并将这几个摘要进行合并。

(4) 合并后，再对合并后的内容根据算法进行提取句子。从而获得了事件的摘要。

### 2.4.5 专题情感计算

专题情感计算反映了网络用户在该专题下讨论过程中的言语倾向于正面或负面的程度。通过专题的情感倾向分析，就可以知道某一网络事件在网民中取得了怎样的反应，以及支持与否定等情感倾向的大致分布情况等问题。一般地，专题的情感分析同时也是舆情预警系统的一个重要参数。文本情感的主要分析步骤如下：

(1) 对 HowNet 的情感词及程度词进行程度标注。

(2) 利用哈尔滨工业大学句法分析技术对文本进行句法分析。

(3) 遍历句法分析结果，找到其中的情感词，以及修饰该情感词的程度词及否定词，其中否定词会考虑距离的影响程度减弱情况，算出该句中单个情感词的极性强度，求和得出整个句子的情感强度。

单个词计算公式为

$$wordsentimentScore = sentimentWordVal \prod_{i=1}^m intense_i \prod_{j=1}^n \frac{ModifiedIntense_j}{dist_j > 2?2:1}$$

式中， $sentenceWordVal$  为情感极性词， $intense_i$  为第  $i$  个程度修饰副词， $ModifiedIntense_j$  为第  $j$  个否定修饰词， $m$  为程度副词修饰的个数， $n$  为否定修饰词的个数， $dist_j$  为第  $j$  个情感，极性词与否定修饰词之间的距离。

(4) 对有情感倾向的句子求和、求平均值，得出单个句子的情感值。

### 2.4.6 专题观点挖掘

一般地,在一个专题下往往存在着多个观点。专题观点是指网友对某一事件或政策的看法。通过分析专题的观点,就可以知道某一个专题内都讨论了哪些问题以及形成了什么样的观点,以及哪些观点引起了网友比较大的争议,而这些观点也是舆情分析者需要格外关注的地方。

专题观点挖掘算法的主要步骤如下:

- (1) 计算专题内所有帖子的内容和其回复与专题本身的相似度。
- (2) 按相似度大小排列句子,并取出不少于 500 条的句子写入文本文档。
- (3) 特征聚类处理与人工检测相结合的方式来分析文本文档中的句子,总结出专题观点及其所占的比率。

此外,由于网络事件的分析涉及自然语言处理以及词句的语义特征分析,而分词方法与策略的好坏将直接决定后继数据分析的准确性与有效性。目前常采用两种方式进行内容的分词处理:一是基于词典的内容分词与分析方法,词典不仅包括了领域的主题词库、词性、情感极性词库,同时包括了停用词的信息;二是采用基于统计学的 N-Gram 分词方法,即通过定义 N 的大小来确定取词的长度,从而对所有取出来的词进行词频统计,其中大于一定阈值的高频词则为系统要自动抽取的词。这两种方式各有优缺点,有时也联合起来使用。但是,采用这两种都无法充分地体现内容中的一些语义特征,如词与词之间存在的上下位、包含等语义关系。因此,一些文献在分词的基础上,通过形成不同的词组结构以及语义上下文来表达出概念词之间存在的语义。在本报告中,对于论坛文本事件的分析,也采用了这一些传统的分词方法;但是对于回复帖内很短的句子,由于缺少上下文的背景语义,利用这种分词的方式很难获得有效的情感与主题识别的关键信息,因此对上述一些算法的实际应用带来新的挑战。这也是本报告未来希望不断深入解决的一个关键问题。

## 2.5 本章小结

通过将社会网络分析方法融入到社会网络舆情的分析过程中,本章构建了核心的分析模型与舆情分析指标体系,从而为本书奠定了一个基础的分析框架。其中,通过对社会网络分析的主要方法与思想介绍,构建了信息发布者模型、信息参与者模型、事件模型以及舆情传播模型等四种核心的分析模型。并在这些分析模型的基础上,建立了数据分析的指标体系,并对这一些指标的物理含义进行说明,同时,对数据分析过程中主要应用的一些算法也进行了定义与分析描述,通过这些工作为后续章节提供了一个有效的分析方法基础。

## 第3章 领域舆情整体网及数据视图

随着社交媒体时代的到来，互联网成为了政府、组织和民众的等表达社情民意的重要平台，这种新的交流方式可以充分体现用户和团体的意愿和态度。从古代的“防民之口甚于防川”，到如今的网络时代“人人都有了自己的麦克风”，舆情的传播发生了重要变化，其重要性也不言而喻。舆情，就是一个风向标。舆情分析，就是针对民众态度的收集和整理，发现相关的意见倾向，客观反映舆情状态。舆情涉及人民生活的方方面面。为了更精确地分析热点，我们针对党的十八大以及十八届三中全会以来，国家在多个领域进行了重点的规划与部署，并针对经济、政治、民生、生态、军事和文化六个核心领域提出了改革政策和方针。在此基础上，结合网络论坛舆情数据的来源特征，从而确定了经济、政治、民生、生态、国际与军事五个领域为研究的核心对象，分别将具体的一个领域中的数据作为整体，来构建领域内基于作者和基于帖子的整体网络，从而研究领域内网民参与热点话题的讨论情况。

### 3.1 整体网络特征分析

整体网研究有助于揭示整个网络的各种结构特征，有利于发现整体网络中的派系聚集的情况，从而揭示出整体系统的整合性，以及整体网的层次性、等级性和阶层性等，从而通过对联系的紧密性进行解释的同时，也实现对整体网成员行为之间的关系分析，找到整体网中联络和分解的模式，以及结构对等的行动者等。

通过第2章社会网络分析方法的描述以及相关模型的建立，本章将在此基础上对整体网络进行构建，同时根据不同的研究对象及其特征，将网络划分为关于主体对象所形成的社会化网络以及客体对象所形成的社会化网络。因此，整体网络分析也分为基于主体对象的社会网络和基于客体对象的社会网络。

#### 3.1.1 构建基于主体对象的社会网络

主体对象是指社会网络中的行动者。构建基于主体对象的某一领域的整体



网络,可以清晰直接地看到论坛用户与该领域中某一主题的对应关系。一般地,主体对象社会网络的构建过程如下:

第一步:确定领域内热点主题,构建分类主题词库。

第二步:使用文本分类方法对领域内原帖进行分类。

第三步:找到每个主题下对应帖子的发帖作者。

第四步:建立发帖作者与具体主题的对应关系。

第五步:通过可视化工具进行可视化。

为了表现出论坛用户随着时间参与不同主题的演化过程,我们将2014年全年数据分为上半年数据和下半年数据,运用上述构建过程来构建社会网络。

### 3.1.2 构建基于客体对象的社会网络

客体对象是指社会网络中的内容实体。构建基于内容的某一领域的整体网络,可以清晰直接地看到论坛帖子与该领域中某一主题的对应关系。一般地,客体对象社会网络的构建过程如下:

第一步:确定领域内热点主题,构建分类主题词库。

第二步:使用分类主题词库对领域内原帖进行相似度比较,抽取出相似度大于一定阈值的原帖与主题。

第三步:建立帖子与具体主题的对应关系。

第四步:通过可视化工具进行可视化。

为了表现出随着时间不同主题中帖子的变化情况,我们把2014年全年数据分为上半年数据和下半年数据,运用上述构建过程来构建社会网络。

## 3.2 经济社区网络特征分析

经济是一个社会的支柱,本章对经济领域整体数据进行了介绍,并对2014年上半年和下半年该领域所有主体和客体在三大主题下的分布情况进行可视化,以发现量的演化。

### 3.2.1 经济领域数据视图

该研究领域的数据中,帖子总数为11989条,评论总数为433528条,参与发帖的总人数为9379人,参与评论的总人数为29802人,具体的分布如表3.1所示。

表 3.1 经济领域数据

主题分类	发帖数/条	评论数/条	发帖人数/个	评论人数/个
股市与股票	1795	227403	1406	5518
房市与房价	3624	44488	2627	2630
货币与汇率	3142	43216	2463	20230
其他	3428	118421	2883	1424
合计	11989	433528	9379	29802

3.2.2 基于主体对象的网络演化分析

由节 3.1.1 中提出的建模过程构建经济领域基于主体对象的社会网络，分别如图 3.1 和图 3.2 所示。

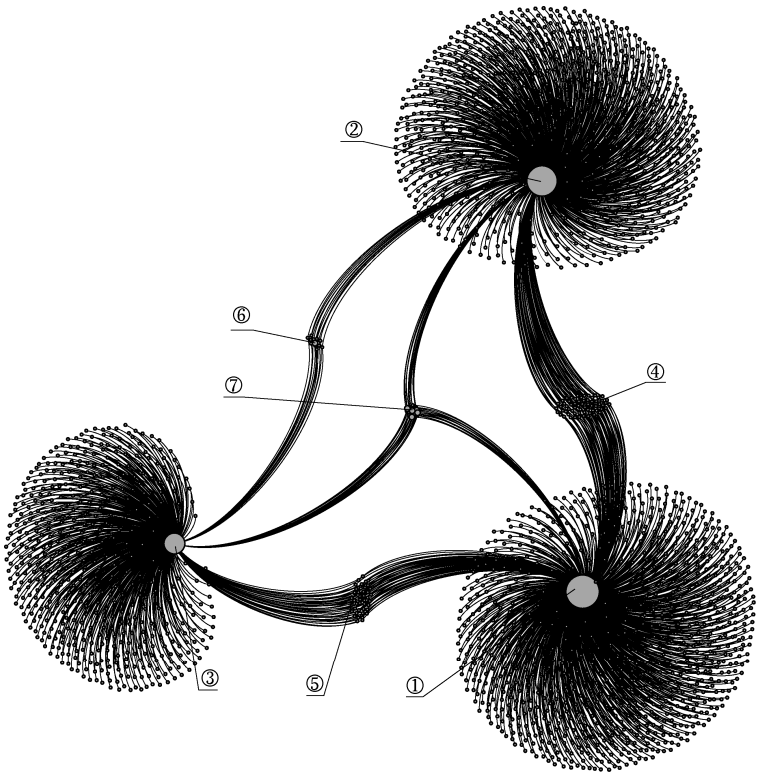


图 3.1 （2014 年上半年）基于人的主题网络（2014.1.1~2014.6.30）

- ① 仅讨论货币与汇率的作者数量为 1236 人，占比为 0.392；
- ② 仅讨论房市与房价的作者数量为 1124 人，占比为 0.359；
- ③ 仅讨论股市与股票的作者数量为 612 人，占比为 0.196；
- ④ 同时讨论货币与汇率、房市与房价的作者数量为 74 人，占比为 0.024；
- ⑤ 同时讨论货币与汇率、股市与股票的作者数量为 63 人，占比为 0.020；
- ⑥ 同时讨论房市与房价、股市与股票的作者数量为 12 人，占比为 0.004；
- ⑦ 同时讨论房市与房价、股市与股票、货币与汇率的作者数量为 16 人，占比为 0.005

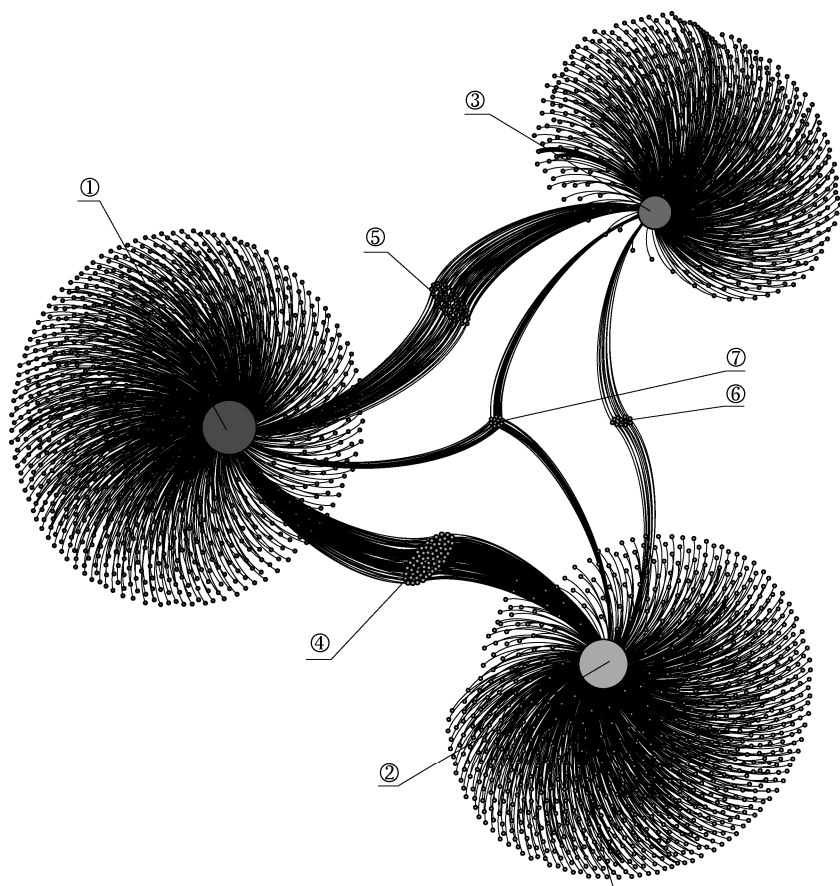


图 3.2 (2014 年下半年) 基于人的主题网络 (2014.7.1~2014.12.31)

- ① 仅讨论货币与汇率的作者数量为 1133 人, 占比为 0.341;
- ② 仅讨论房市与房价的作者数量为 1335 人, 占比为 0.403;
- ③ 仅讨论股市与股票的作者数量为 669 人, 占比为 0.202;
- ④ 同时讨论货币与汇率、房市与房价的作者数量为 109 人, 占比为 0.033;
- ⑤ 同时讨论货币与汇率、股市与股票的作者数量为 45 人, 占比为 0.015;
- ⑥ 同时讨论房市与房价、股市与股票的作者数量为 6 人, 占比为 0.002;
- ⑦ 同时讨论房市与房价、股市与股票、货币与汇率的作者数量为 16 人, 占比为 0.004

图 3.1 是根据 2014 年上半年的数据构建的论坛中较活跃的用户参与经济领域中热点主题的社会网络图。从中可以发现, 讨论房市与房价、货币与汇率两个主题的活跃用户多于讨论股市与股票主题的用户, 只有少数用户同时活跃在经济领域的三大主题内。我们可以将用户分为三大类: 第一类用户参与三个主题的讨论; 第二类用户参与两个主题的讨论; 第三类用户仅仅参与其中一个主题的讨论。

第一类用户的总人数最少，他们参与了经济领域所有热点主题的讨论，这部分人是经济领域的中流砥柱，最有可能成为经济领域的意见领袖。第二类用户是经济领域的重要参与者，他们经常在网上参与某些热点话题的讨论。第三类用户可能单纯对某一热点话题感兴趣，这种用户最有可能发生变动，因为他们参与某一话题最大的因素是阶段性的需求。例如，想要买房子或出手房产的人在某个时期会关注房产与房价信息。但是，也有一部分人可能是经济领域某一主题的长期参与者，如长期投资股票的用户可能关注更多的是股票相关信息。

图 3.2 是根据 2014 年下半年的数据构建的论坛中较活跃的用户参与经济领域中热点主题的社会网络图。从中可以发现，经济领域的参与用户总体数量维持基本不变，各主题所占比例大体稳定，只有房市与房价人数占比有所提高、货币与汇率人数占比有所降低。房市与房价主题所占比例的提高，可能是因为下半年国家对与房地产的相关政策引起了广大网友的关注，将一部分网民从货币与汇率主题吸引到房市与房价主题。

将图 3.1 与图 3.2 中参与经济领域讨论网民的占比绘图得图 3.3。从图 3.3 中明显看出，关注房市与房价的作者所占比例在下半年有明显上升；关注货币与汇率的作者所占比例在下半年有明显下降；大部分作者仅关注一个主题，说明广大网民的关注点较为单一；关注两个甚至三个主题的作者很有可能成为经济领域的意见领袖，他们的言论将对其他网民产生较大的影响。

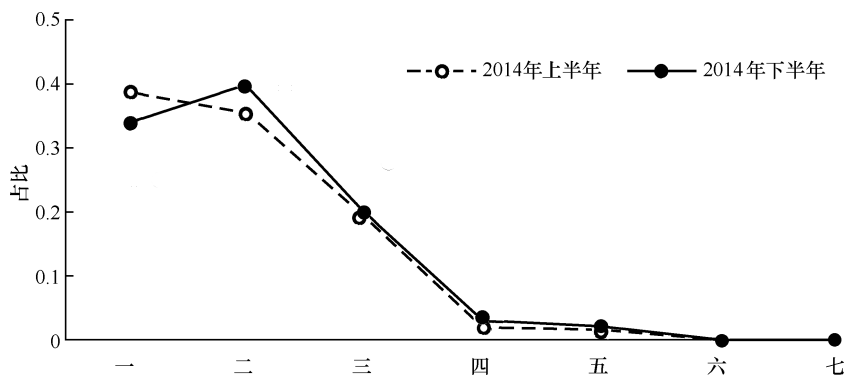


图 3.3 参与各主题的用户占比

- 一、货币与汇率讨论参考者；二、房市与房价讨论参考者；三、股市与股票讨论参考者；  
 四、房市与房价、货币与汇率讨论参考者；五、股市与股票、货币与汇率讨论参考者；  
 六、股市与股票、房市与房价讨论参考者；七、股市与股票、房市与房价、货币与汇率讨论参考者

基于主体对象的社会网络基本指标如表 3.2 所示。2014 年上半年和 2014 年下半年的指标基本相似。平均度上半年略大于下半年，说明上半年的网民“综合度”更大，即上半年网民的帖子平均涉及的主题更多；网络直径都为 4，说明任意两个节点之间最大的距离为 4；图密度都为 0.001 说明图 3.1 与图 3.2 都

是稀疏图；模块化指标分别为 0.569 和 0.592，说明图 3.2 划分的社区质量更高一些；平均路径长度为 3.179 和 3.217，下半年的平均长度稍大说明下半年中的参与作者略多，与图 3.2 比图 3.1 反映出参与的网民数量稍多一致。

表 3.2 网络基本指标对比

调研时间	平均度	网络直径	图密度	模块化	平均路径长度
2014 年上半年	2.147	4	0.001	0.569	3.179
2014 年下半年	2.125	4	0.001	0.592	3.217

### 3.2.3 基于客体对象的网络演化分析

由节 3.1.2 中提出的建模过程构建经济领域基于客体对象的社会网络分别如图 3.4 和图 3.5 所示。

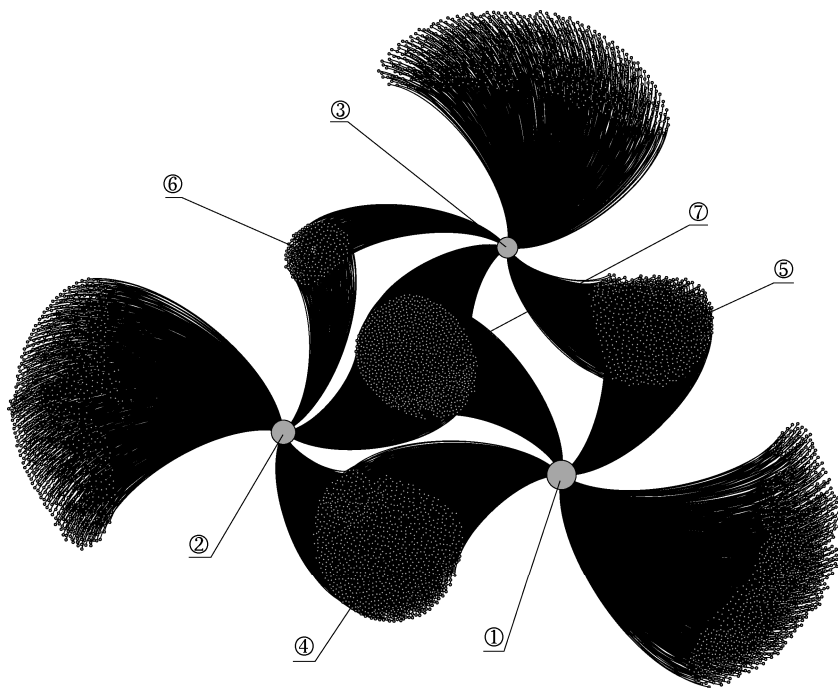


图 3.4 （2014 年上半年）基于内容的主题网络（2014.1.1~2014.6.30）

- ① 仅讨论货币与汇率的帖子数量为 967 个，占比为 0.189；
- ② 仅讨论房市与房价的帖子数量为 765 个，占比为 0.146；
- ③ 仅讨论股市与股票的帖子数量为 648 个，占比为 0.126；
- ④ 同时讨论货币与汇率、房市与房价的帖子数量为 986 个，占比为 0.193；
- ⑤ 同时讨论货币与汇率、股市与股票的帖子数量为 642 个，占比为 0.125；
- ⑥ 同时讨论房市与房价、股市与股票的帖子数量为 179 个，占比为 0.036；
- ⑦ 同时讨论房市与房价、股市与股票、货币与汇率的帖子数量为 947 人，占比为 0.185

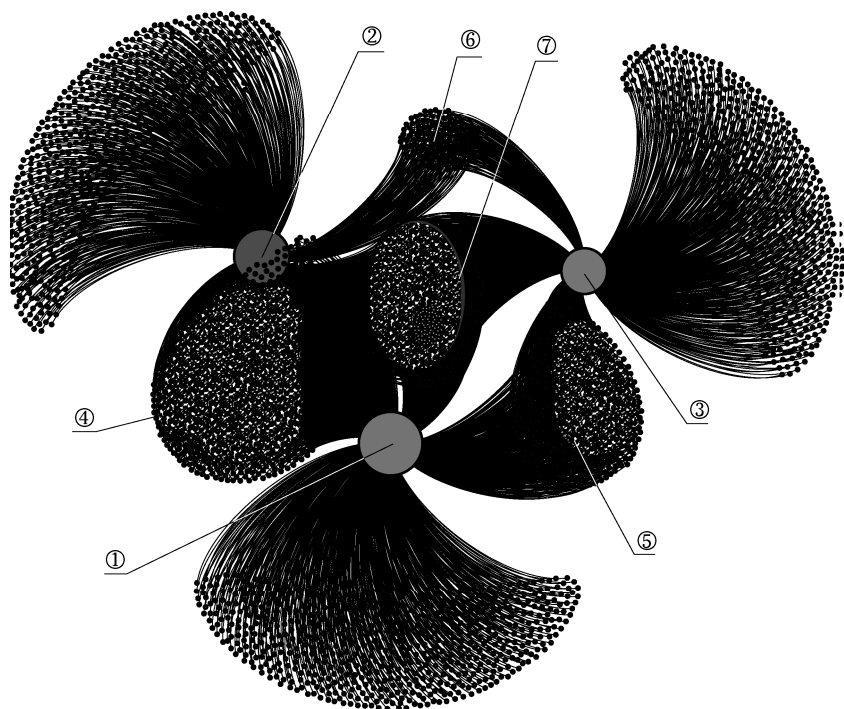


图 3.5 基于帖子的主题网络（2014.7.1~2014.12.31）

- ① 仅讨论货币与汇率的帖子数量为 608 个，占比为 0.168；
- ② 仅讨论房市与房价的帖子数量为 612 个，占比为 0.169；
- ③ 仅讨论股市与股票的帖子数量为 518 个，占比为 0.143；
- ④ 同时讨论货币与汇率、房市与房价的帖子数量为 727 个，占比为 0.202；
- ⑤ 同时讨论货币与汇率、股市与股票的帖子数量为 449 个，占比为 0.124；
- ⑥ 同时讨论房市与房价、股市与股票的帖子数量为 122 个，占比为 0.034；
- ⑦ 同时讨论房市与房价、股市与股票、货币与汇率的帖子数量为 578 个，占比为 0.160

图 3.4 是根据 2014 年上半年的数据构建的经济领域中热点主题的社会网络图。可以发现，单独讨论股市与股票、房市与房价、货币与汇率三个主题的帖子数量大致相当，货币与汇率的数量略大于其他两个主题，说明网民对于股市与股票、房市与房价、货币与汇率这三个主题的关注度差别不太大。同时可以发现，论坛中的帖子具有很强的综合倾向，即一个帖子讨论的内容涉及两个甚至三个主题。同时涉及货币与汇率和股市与股票两个主题的帖子数量较多，反映出这两个主题之间具有较强的关联性；同时涉及股市与股票、房市与房价、货币与汇率的帖子数量也较多，反映出部分网民对于经济领域的综合了解，这些网民可能会成为经济领域的意见领袖。

图 3.5 是根据 2014 年下半年的数据构建的经济领域中热点主题的社会网络图。可以发现，讨论股市与股票、房市与房价、货币与汇率三个主题的帖子数量少于上半年的帖子数量，说明下半年网民对于经济领域关注的热度总体有所下降；但帖子所属的主题占比分布也与上半年相似，仅关注房市与房价的帖子数量占比较上半年有所提高，说明网民对房市与房价专题更为关心；同时涉及股市与股票、房市与房价、货币与汇率的帖子数量仍然较多。

将图 3.4 与图 3.5 中参与经济领域讨论网民的占比绘图得图 3.6。从图 3.6 中明显看到，关注房市与房价、股市与股价的帖子占比在下半年有明显上升；关注货币与汇率的帖子占比在下半年有明显下降；同时参与两个主题讨论的作者比例变化比较小；同时参与三个主题的作者比例明显上升。

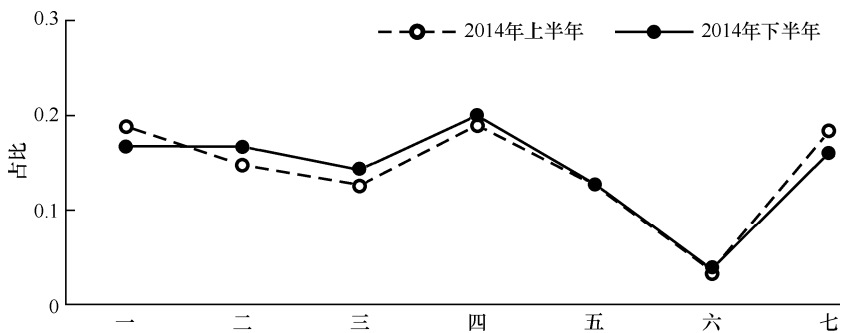


图 3.6 各主题的帖子占比

- 一、货币与汇率讨论参考者；二、房市与房价讨论参考者；三、股市与股票讨论参考者；  
四、房市与房价、货币与汇率讨论参考者；五、股市与股票、货币与汇率讨论参考者；  
六、股市与股票、房市与房价讨论参考者；七、股市与股票、房市与房价、货币与汇率讨论参考者

基于客体对象的社会网络基本指标如表 3.3 所示。2014 年上半年和 2014 年下半年的指标基本相似。平均度上半年略大于下半年，说明上半年的帖子“综合度”更大一点，即上半年帖子平均描述的主题更多；网络直径都为 4，说明任意两个节点间最大的距离为 4；图密度都为 0.001 说明图 3.4 和图 3.5 都是稀疏图；模块化指标分别为 0.246 和 0.261，总体上是较低的指标值，说明这两个网络中的社区不明显，不易被发现；平均路径长度为 2.480 和 2.529，下半年的平均长度较大说明下半年的帖子数增多。

表 3.3 网络基本指标对比

调研时间	平均度	网络直径	图密度	模块化	平均路径长度
2014 年上半年	3.440	4	0.001	0.246	2.480
2014 年下半年	3.355	4	0.001	0.261	2.529

### 3.3 政治社区网络特征分析

政治是一个国家的体制，本节对政治领域整体数据进行了介绍，并对 2014 年上半年和下半年该领域所有主体和客体在三大主题下的分布情况进行了可视化，以发现量的演化。

#### 3.3.1 政治领域数据视图

该研究领域的数据中，帖子总数为 9560 条，评论总数为 172906 条，参与发帖的总人数为 5774 人，参与评论的总人数为 63154 人，具体的分布如表 3.4 所示。

表 3.4 政治领域数据视图

主题分类	发帖数/条	评论数/条	发帖人数/个	评论人数/个
国企改革	3232	47071	1994	17856
依法治国	3930	95713	2293	32004
反腐风暴	2398	30122	1487	13294
合计	9560	172906	5774	63154

#### 3.3.2 基于主体对象的网络演化分析

由节 3.1.1 中提出的建模过程构建政治领域基于主体对象的社会网络分别如图 3.7 和图 3.8 所示。

图 3.7 是根据 2014 年上半年的数据构建的论坛中较活跃的用户参与政治领域中热点主题的社会网络图。可以发现，讨论国企改革、反腐风暴两个主题的活跃用户多于讨论依法治国主题的用户，只有少量用户同时活跃在政治领域的三大主题下我们可以将用户分为三大类：第一类用户同时参与三个主题的讨论；第二类用户参与两个主题的讨论；第三类用户仅仅参与其中一個主题的讨论。

第一类用户的总人数最少，他们参与了政治领域所有热点主题的讨论，这部分人是政治领域的中流砥柱，最有可能成为政治领域的意见领袖。第二类用户是政治领域的重要参与者，他们经常在网上参与某些热点话题的讨论。第三类用户可能单纯对某一热点话题感兴趣，这种用户最有可能发生变动，因为他们参与某一话题最大的因素是阶段性的需求。



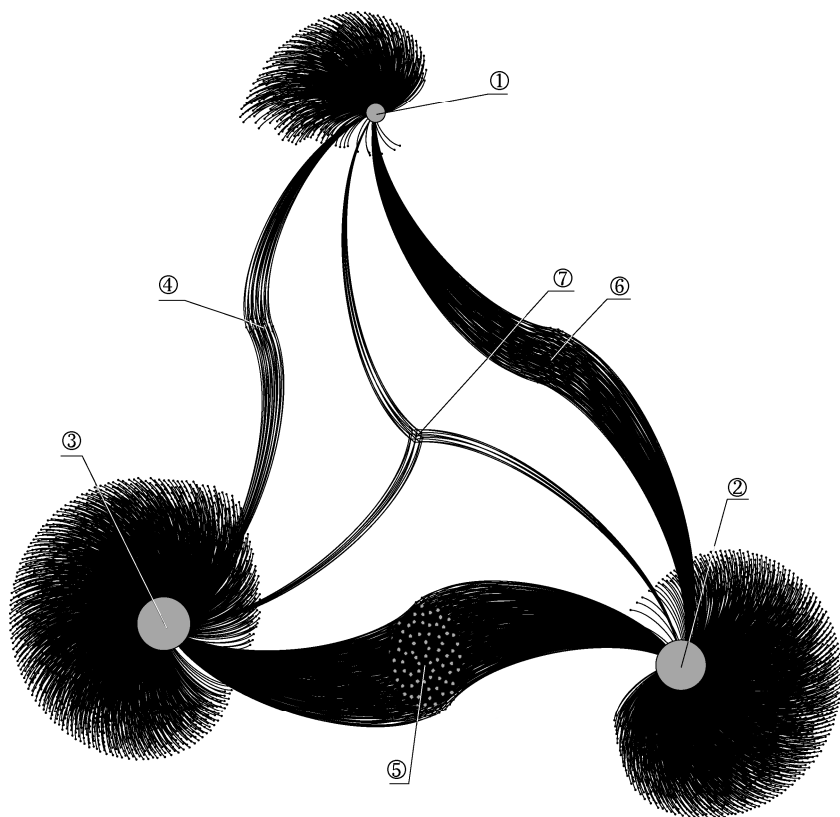


图 3.7 (2014 年上半年) 基于人的主题网络 (2014.1.1~2014.6.30)

- ① 仅讨论依法治国的作者数量为 402 人, 占比为 0.146;
- ② 仅讨论反腐风暴的作者数量为 933 人, 占比为 0.314;
- ③ 仅讨论办国企改革的作者数量为 1394 人, 占比为 0.451;
- ④ 同时讨论依法治国、国企改革的作者数量为 49 人, 占比为 0.016;
- ⑤ 同时讨论国企改革、反腐风暴的作者数量为 150 人, 占比为 0.051;
- ⑥ 同时讨论反腐风暴、依法治国的作者数量为 28 人, 占比为 0.009;
- ⑦ 同时讨论反腐风暴、依法治国、国企改革的作者数量为 36 人, 占比为 0.013

图 3.8 是根据 2014 年下半年的数据构建的论坛中较活跃的用户参与政治领域中热点主题的社会网络图。可以发现, 政治领域的参与用户总体数量有明显上升, 其国企改革专题和反腐风暴专题热度维持稳定, 依法治国专题被讨论的热度明显增加, 显然与国家的相关政策有着密切的关系, 如《中共中央关于全面推进依法治国若干重大问题的决定》的发布等。

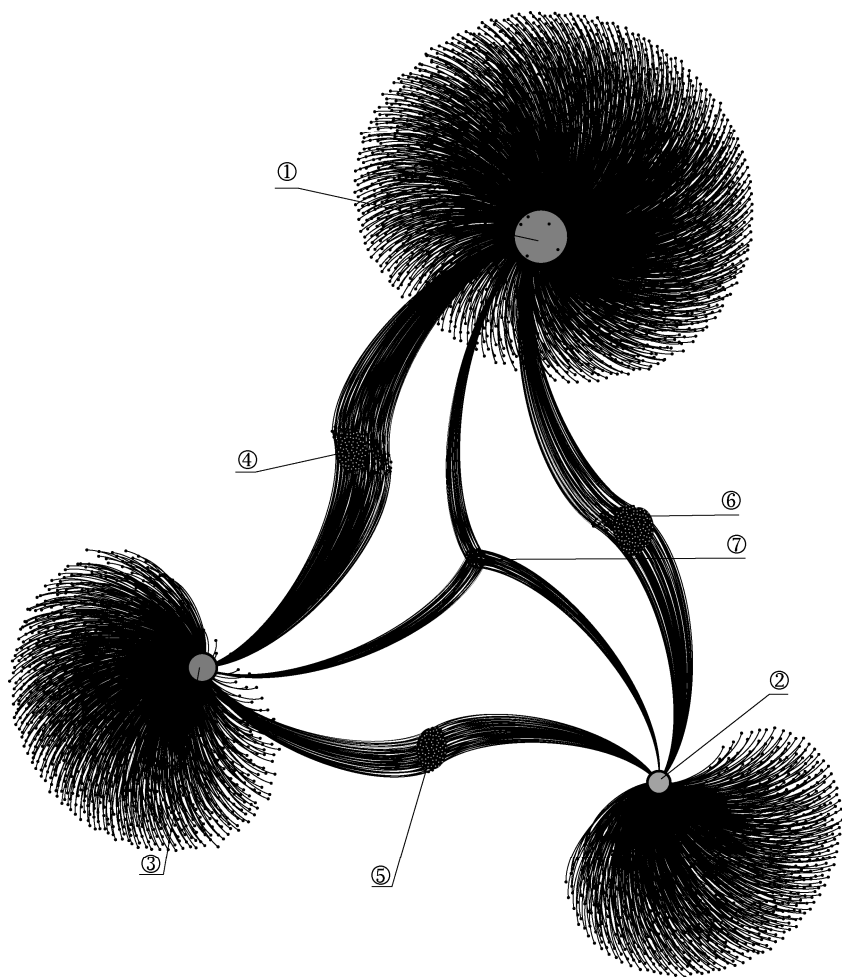


图 3.8 (2014 年下半年) 基于人的主题网络 (2014.7.1~2014.12.31)

- ① 仅讨论依法治国的作者数量为 3251 人，占比为 0.546；
- ② 仅讨论反腐风暴的作者数量为 992 人，占比为 0.167；
- ③ 仅讨论办国企改革的作者数量为 1313 人，占比为 0.211；
- ④ 同时讨论依法治国、国企改革的作者数量为 134 人，占比为 0.023；
- ⑤ 同时讨论国企改革、反腐风暴的作者数量为 99 人，占比为 0.016；
- ⑥ 同时讨论反腐风暴、依法治国的作者数量为 108 人，占比为 0.017；
- ⑦ 同时讨论反腐风暴、依法治国、国企改革的作者数量为 57 人，占比为 0.010

将图 3.7 与图 3.8 中参与政治领域讨论网民的占比绘图得图 3.9。从图 3.9 中明显看到，关注反腐风暴专题、国企改革专题的作者所占比例在下半年有明显下降，关注依法治国专题的作者数量明显上升。同时可以看出，大部分作者仅关注一个主题，说明广大网民的关注点较为单一；关注两个甚至三个主题的

作者很有可能成为政治领域的意见领袖，他们的言论将对其他网民产生较大的影响。

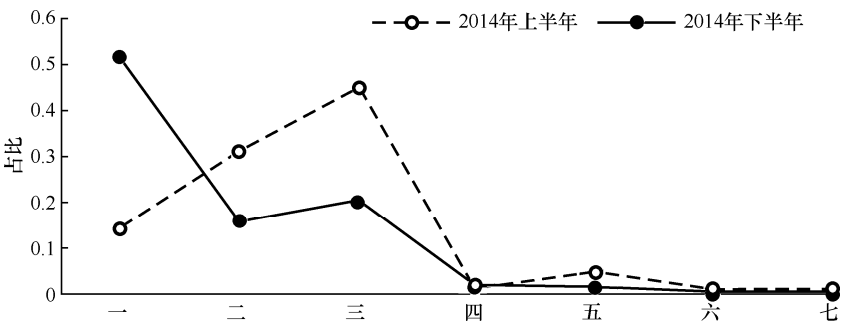


图 3.9 参与各主题的用户占比

一、依法治国讨论参考者；二、反腐风暴讨论参考者；三、国企改革讨论参考者；  
四、依法治国、国企改革讨论参考者；五、国企改革、反腐风暴讨论参考者；  
六、反腐风暴、依法治国讨论参考者；七、反腐风暴、依法治国、国企改革讨论参考者

基于主体对象的社会网络基本指标如表 3.5 所示。2014 年上半年和 2014 年下半年的指标基本相似。平均度下半年略大于上半年，说明下半年的网民“综合度”更大，即下半年网民平均涉及的主题更多；网络直径都为 4，说明任意两个节点之间最大的距离为 4；图密度都为 0.001 说明图 3.7 和图 3.8 都是稀疏图；模块化指标分别为 0.545 和 0.523，说明上半年划分的社区质量更高；平均路径长度为 3.115 和 3.091，下半年的平均长度较小说明下半年的作者所涉及的平均主题略高于上半年。

表 3.5 网络基本指标对比

调研时间	平均度	网络直径	图密度	模块化	平均路径长度
2014 年上半年	2.136	4	0.001	0.545	3.115
2014 年下半年	2.180	4	0.001	0.523	3.091

3.3.3 基于客体对象的网络演化分析

由节 3.1.2 中提出的建模过程构建政治领域基于客体对象的社会网络分别如图 3.10 和图 3.11 所示。

图 3.10 是根据 2014 年上半年的数据构建的政治领域中热点主题的社会网络图。可以发现，单独讨论依法治国专题帖子的数量较少，讨论国企改革专题与反腐风暴专题的帖子数量大致相当，说明 2014 年上半年网民对国企改革和反腐风暴的关注度略大于依法治国专题。同时可以发现，涉及国企改革专题与反腐风暴的关注度略大于依法治国专题。

腐风暴专题的帖子数量比较可观，说明这两个专题具有较强的关联性；同时涉及依法治国和反腐风暴两个主题的帖子数量次之；同时涉及依法治国和国企改革两个主题的帖子数量最少，说明两个专题的关系较弱。

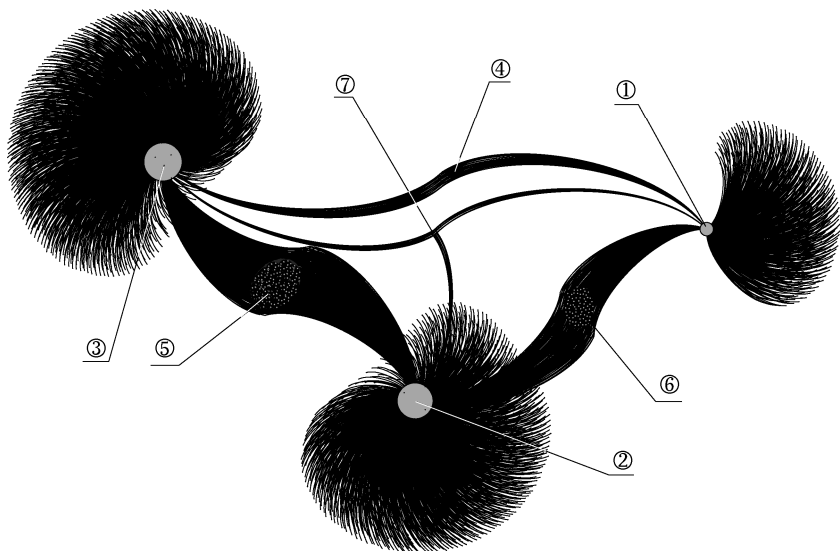


图 3.10 (2014 年上半年) 基于内容的主题网络 (2014.1.1~2014.6.30)

- ① 仅讨论依法治国的帖子数量为 428 个，占比为 0.150；
- ② 仅讨论反腐风暴的帖子数量为 933 个，占比为 0.328；
- ③ 仅讨论办国企改革的帖子数量为 1106 个，占比为 0.388；
- ④ 同时讨论依法治国、国企改革帖子数量为 16 个，占比为 0.006；
- ⑤ 同时讨论国企改革、反腐风暴的帖子数量为 275 个，占比为 0.097；
- ⑥ 同时讨论反腐风暴、依法治国的帖子数量为 83 个，占比为 0.029；
- ⑦ 同时讨论反腐风暴、依法治国、国企改革的帖子数量为 7 个，占比为 0.002

图 3.11 是根据 2014 年下半年的数据构建的政治领域中热点主题的社会网络图。可以发现，讨论反腐风暴专题与国企改革专题的帖子数量维持稳定，讨论依法治国专题的帖子数量快速上升，同时影响到涉及依法治国和反腐风暴两个专题的帖子数量增多、涉及依法治国和国企改革两个专题的帖子数量增多。

将图 3.10 与图 3.11 中参与政治领域讨论帖子的占比绘图得图 3.12。从图 3.12 中明显看到，关注国企改革、反腐风暴的帖子所占比例在下半年有明显下降，这是由于下半年关于依法治国的帖子数量迅速上升，导致数量维持稳定的国企改革专题和反腐风暴专题所占比例下降，这个结果和参与作者的比例图相一致。

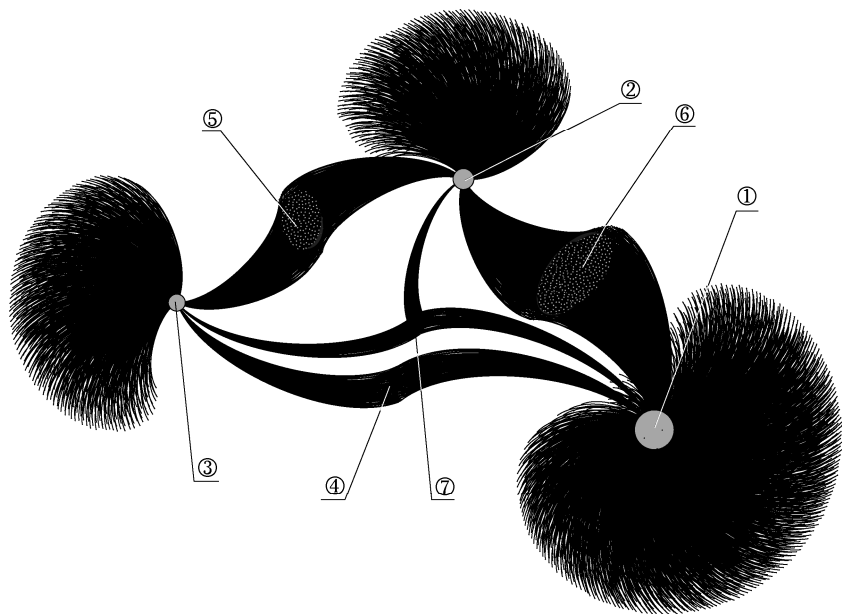


图 3.11 (2014 年下半年) 基于内容的主题网络 (2014.7.1~2014.12.31)

- ① 仅讨论依法治国的帖子数量为 2764 个, 占比为 0.441;
- ② 仅讨论反腐风暴的帖子数量为 1186 个, 占比为 0.189;
- ③ 仅讨论办国企改革帖子数量为 1265 个, 占比为 0.202;
- ④ 同时讨论依法治国、国企改革的帖子数量为 114 个, 占比为 0.018;
- ⑤ 同时讨论国企改革、反腐风暴的帖子数量为 271 个, 占比为 0.043;
- ⑥ 同时讨论反腐风暴、依法治国的帖子数量为 608 个, 占比为 0.097;
- ⑦ 同时讨论反腐风暴、依法治国、国企改革的帖子数量为 67 个, 占比为 0.010

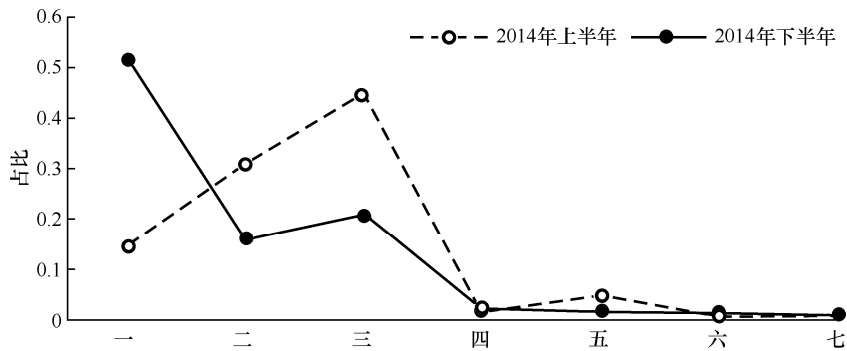


图 3.12 各主题的所属帖子占比

一、依法治国讨论参考者；二、反腐风暴讨论参考者；三、国企改革讨论参考者；  
四、依法治国、国企改革讨论参考者；五、国企改革、反腐风暴讨论参考者；  
六、反腐风暴、依法治国讨论参考者；七、反腐风暴、依法治国、国企改革讨论参考者

基于客体对象的社会网络基本指标如表 3.6 所示。2014 年上半年和 2014 年下半年的指标基本相似。平均度下半年略大于上半年，说明下半年的帖子“综合度”更大一点，即下半年帖子平均描述的主题更多；网络直径都为 4，说明任意两个节点间最大的距离为 4；图密度都为 0.001 说明图 3.10 和图 3.11 都是稀疏图；模块化指标分别为 0.518 和 0.487，较低的指标值说明这两个网络中的社区不明显，不易被发现；平均路径长度为 3.051 和 2.955，上半年的平均路径长度较小说明上半年任意两点的最短路径长度更大，即上半年的数据更为“分散”。

表 3.6 网络基本指标对比

调研时间	平均度	网络直径	图密度	模块化	平均路径长度
2014 年上半年	2.240	4	0.001	0.518	3.051
2014 年下半年	2.368	4	0.001	0.487	2.955

### 3.4 民生社区网络特征分析

邓小平同志说：“社会主义财富属于人民，社会主义的致富是全民共同致富。”构建和谐社会，就是要把民生问题作为重中之重，让广大人民群众有活干，有学上，有饭吃，有衣穿，有屋住，病有医，老有养，生活幸福，都过上好日子。民生的主要问题总结起来一共六个方面，即教育、就业、收入分配、社会保障、医疗卫生和社会管理。本节中将民生领域总结为教育与就业专题、收入分配专题以及社会与医疗保险专题共三大主题，分别构建这三大主题的特征词库，对原始数据进行分类，根据分类结果构建社区网络并进行分析。

#### 3.4.1 民生领域数据视图

本研究领域的数据中，帖子总数为 13031 条，评论总数为 323047 条，参与发帖的总人数为 10133 人，参与评论的总人数为 72608 人，具体的分布如表 3.7 所示。

表 3.7 民生领域数据

主题分类	发帖数/条	评论数/条	发帖人数/个	评论人数/个
教育与就业	3560	74621	2798	16385
收入分配	2761	95606	2073	24400
社会与医疗保险	3342	75261	2368	18922
其他	3368	77570	2894	12901
合计	13031	323047	10133	72608

### 3.4.2 基于主体对象的网络演化分析

由节 3.1.1 中提出的建模过程构建民生领域基于主体对象的社会网络分别如图 3.13 和图 3.14 所示。

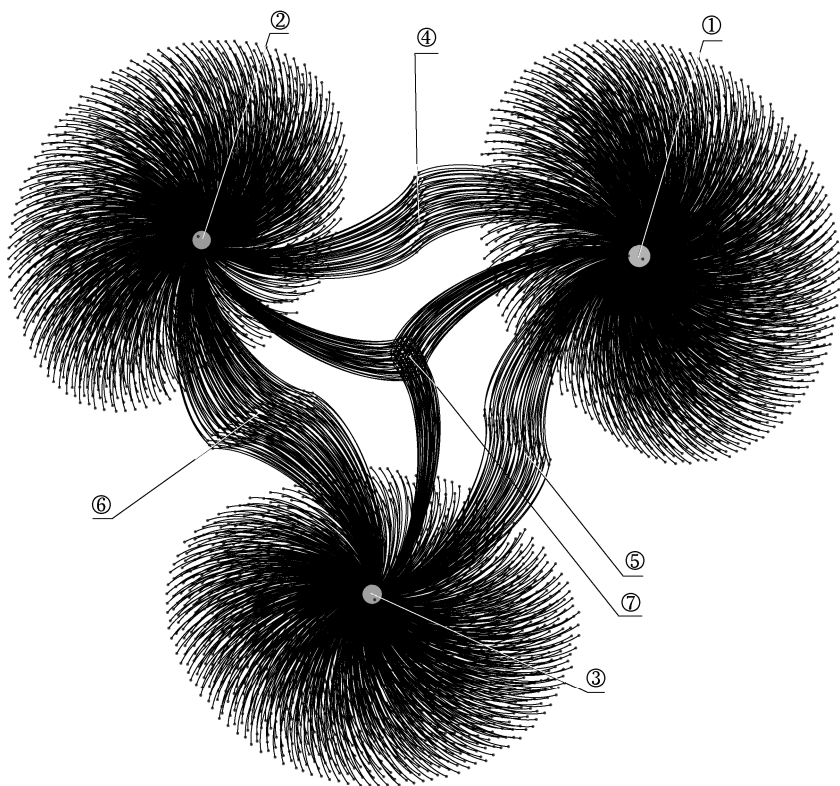


图 3.13 (2014 年上半年) 基于人的主题网络 (2014.1.1~2014.6.30)

- ① 仅讨论教育与就业的作者数量为 1046 人, 占比为 0.394;
- ② 仅讨论收入分配的作者数量为 694 人, 占比为 0.262;
- ③ 仅讨论社会与医疗保险的作者数量为 821 人, 占比为 0.310;
- ④ 同时讨论教育与就业、收入分配的作者数量为 19 人, 占比为 0.007;
- ⑤ 同时讨论教育与就业、社会与医疗保险的作者数量为 22 人, 占比为 0.008;
- ⑥ 同时讨论社会与医疗保险、收入分配的作者数量为 35 人, 占比为 0.013;
- ⑦ 同时讨论教育与就业、收入分配、社会与医疗保险的作者数量为 17 人, 占比为 0.006

图 3.13 是根据 2014 年上半年数据构建论坛中的用户参与民生领域中热点主题的社会网络图。可以发现, 讨论教育与就业主题的参与人数最多, 讨论社

会与医疗保险主题的参与人次次之，讨论收入分配主题的参与人数最少。绝大多数用户都只是发表了单一主题相关的帖子，极少部分用户发表的帖子涉及了两个及三个主题。在这部分用户中，同时参与社会与医疗保险和收入分配的用户数量最多，这有可能是这两个主题的相关性较强的原因。

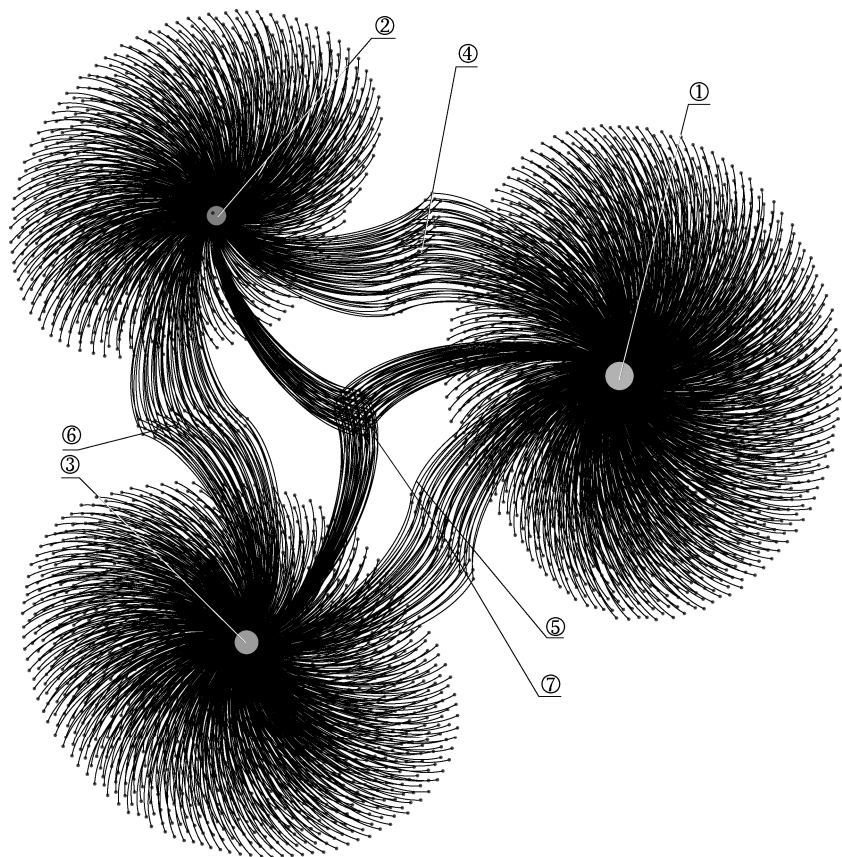


图 3.14 (2014 年下半年) 基于人的主题网络 (2014.7.1~2014.12.31)

- ① 仅讨论教育与就业的作者数量为 958 人，占比为 0.413；
- ② 仅讨论收入分配的作者数量为 502 人，占比为 0.216；
- ③ 仅讨论社会与医疗保险的作者数量为 778 人，占比为 0.335；
- ④ 同时讨论教育与就业、收入分配的作者数量为 23 人，占比为 0.010；
- ⑤ 同时讨论教育与就业、社会与医疗保险的作者数量为 21 人，占比为 0.009；
- ⑥ 同时讨论社会与医疗保险、收入分配的作者数量为 23 人，占比为 0.010；
- ⑦ 同时讨论教育与就业、收入分配、社会与医疗保险的作者数量为 15 人，占比为 0.007

图 3.14 是根据 2014 年下半年数据构建论坛中的用户参与民生领域中热点主题的社会网络图。可以发现，民生领域的参与用户总体数量变化不太明显，



但是各主题所占比例有所变化。教育与就业主题和社会与医疗保险主题形成的社区较大，参与作者比较多；收入分配形成的社区较小，参与者数量比较少；参与两个及三个主题的作者数量有所增加，这部分作者的增加是作者深入参与民生领域舆情的演化结果。

将图 3.13 与图 3.14 中参与民生领域讨论帖子的占比绘图得图 3.15。从图 3.15 中可以看到，关注教育与就业和社会与医疗保险的作者所占比例在下半年有所上升，关注收入分配的作者所占比例在下半年有所下降。同时可以看出，大部分作者仅关注一个主题，说明广大网民的关注点较为单一。关注两个甚至三个主题的作者相对较少，这部分用户发表了不只一篇帖子，是民生领域的活跃用户。

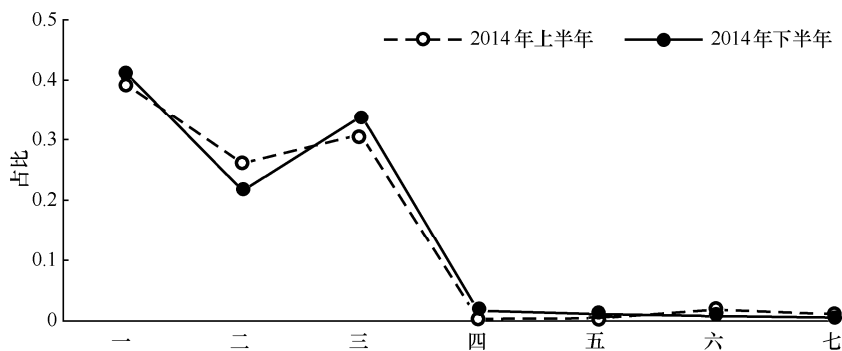


图 3.15 参与各主题的用户占比

一、教育与就业讨论参考者；二、收入分配讨论参考者；三、社会与医疗保障讨论参考者；  
四、教育与就业、收入分配讨论参考者；五、教育与就业、社会与医疗讨论参考者；  
六、收入分配、社会与医疗保障讨论参考者；七、教育与就业、收入分配、社会与医疗保障讨论参考者

基于主体对象的社会网络基本指标如表 3.8 所示。2014 年上半年和 2014 年下半年的指标基本相似。平均度上半年略大于下半年，说明上半年的网民“综合度”更大，即上半年网民的帖子平均涉及的主题更多；网络直径都为 4，说明任意两个节点之间最大的距离为 4；图密度都为 0.001，说明图 3.13 和图 3.14 都是稀疏图；模块化指标分别为 0.598 和 0.595，说明图 3.13 划分的社区质量更高，这也说明上半年同属于两个或三个以上社区的作者相对较少；平均路径长度分别为 3.23 和 3.216，基本相似。

表 3.8 网络基本指标对比

调研时间	平均度	网络直径	图密度	模块化	平均路径长度
2014 年上半年	2.144	4	0.001	0.598	3.230
2014 年下半年	2.125	4	0.001	0.595	3.216

3.4.3 基于客体对象的网络演化分析

由节 3.1.2 中提出的建模过程构建民生领域基于客体对象的社会网络分别如图 3.16 和图 3.17 所示。

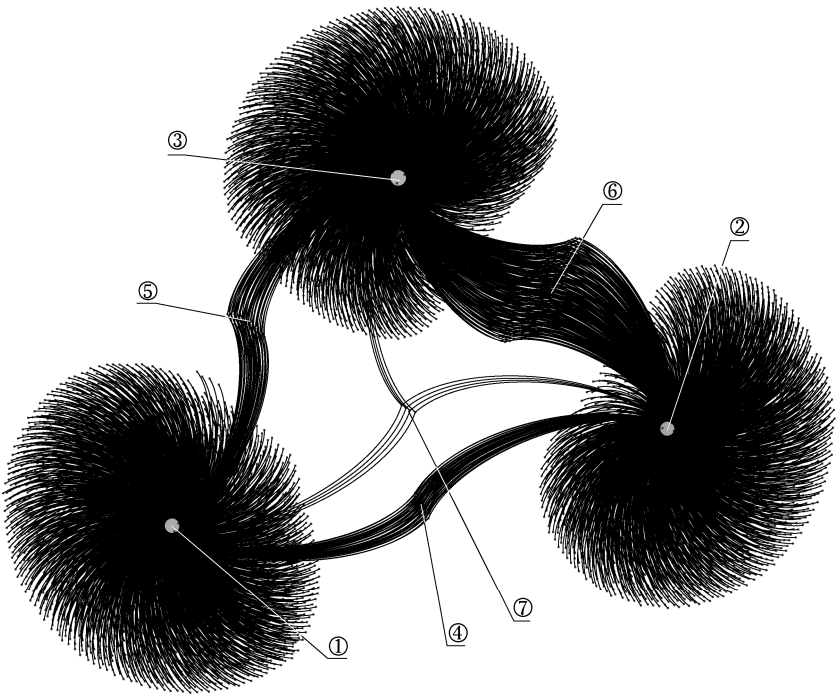


图 3.16 （2014 年上半年）基于内容的主题网络（2014.1.1~2014.6.30）

- ① 仅讨论教育与就业的帖子数量为 1819 个，占比为 0.330；
- ② 仅讨论收入分配的帖子数量为 1592 个，占比为 0.290；
- ③ 仅讨论社会与医疗保险的帖子数量为 1788 个，占比为 0.326；
- ④ 同时讨论教育与就业、收入分配的帖子数量为 36 个，占比为 0.007；
- ⑤ 同时讨论教育与就业、社会与医疗保险的帖子数量为 46 个，占比为 0.008；
- ⑥ 同时讨论社会与医疗保险、收入分配的帖子数量为 207 个，占比为 0.038；
- ⑦ 同时讨论教育与就业、收入分配、社会与医疗保险的帖子数量为 4 个，占比为 0.001

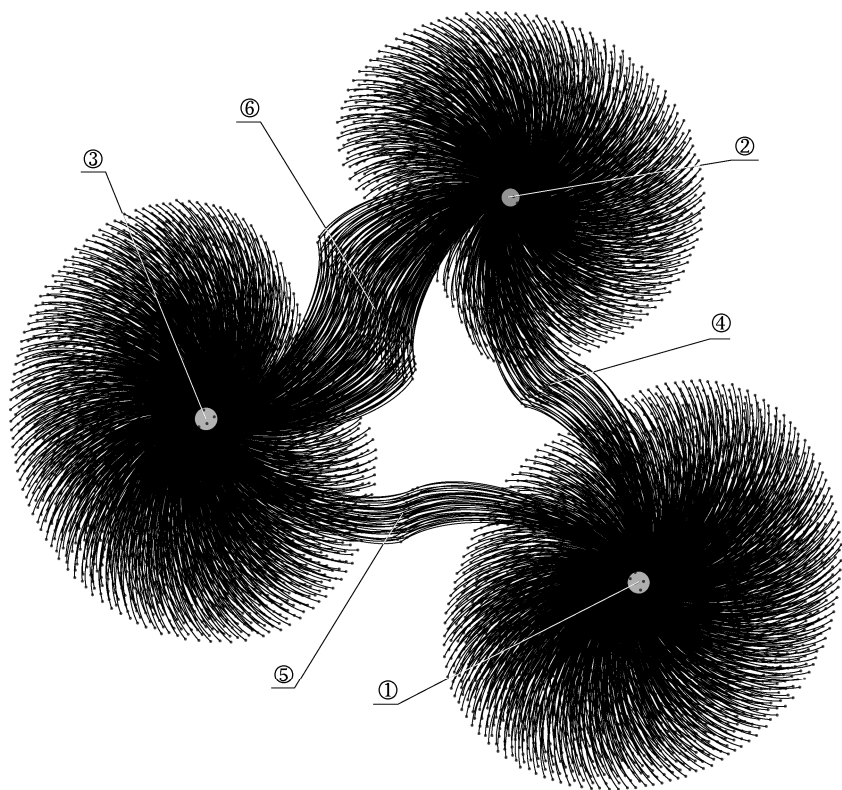


图 3.17 (2014 年下半年) 基于内容的主题网络 (2014.7.1~2014.12.31)

- ① 仅讨论教育与就业的帖子数量为 1746 个, 占比为 0.352;
- ② 仅讨论收入分配的帖子数量为 1288 个, 占比为 0.258;
- ③ 仅讨论社会与医疗保险的帖子数量为 1687 个, 占比为 0.340;
- ④ 同时讨论教育与就业、收入分配的帖子数量为 46 个, 占比为 0.009;
- ⑤ 同时讨论教育与就业、社会与医疗保险的帖子数量为 39 个, 占比为 0.009;
- ⑥ 同时讨论社会与医疗保险、收入分配的帖子数量为 154 个, 占比为 0.032;
- ⑦ 同时讨论教育与就业、收入分配、社会与医疗保险的帖子数量为 0 个, 占比为 0, 因此图中无显示此标注

图 3.16 是根据 2014 年上半年数据构建民生领域中热点主题的社会网络图。可以发现, 单独讨论教育与就业主题的帖子数最多, 其次是社会与医疗主题, 最后是收入分配主题, 说明网民对于教育与就业问题的关注度较大、其次是社会与医疗问题、最后是收入分配问题。同时可以发现, 论坛中帖子的综合倾向比较弱, 即一个帖子讨论的内容涉及两个甚至三个主题的情况相对较少。同时涉及社会与医疗保险和收入分配这两个主题的帖子数量较多, 反映出这两个主题之间具有较强的关联性; 同时涉及三个主题的帖子极少, 说明这三个主题的

相关性不是很强，侧面反映了民生领域涉及问题的宽度较广。

图 3.17 是根据 2014 年下半年数据构建民生领域中热点主题的社会网络图。可以发现，单独讨论各个主题的帖子数量排名上，仍然是教育与就业主题的帖子数最多、其次是社会与医疗保险主题，最后是收入分配主题，说明网民对于教育与就业问题的关注度较大、其次是社会与医疗问题、最后是收入分配问题。同时可以发现，论坛中帖子的综合倾向比较弱，即一个帖子讨论的内容涉及两个甚至三个主题的情况相对较少。同时涉及社会与医疗保险和收入分配这两个主题的帖子数量较多，反映出这两个主题之间具有较强的关联性；没有同时涉及三个主题的帖子，说明这三个主题之间没有相关性，侧面反映了民生领域涉及问题的宽度很广。

将图 3.16 与图 3.17 中参与民生领域讨论帖子的占比绘图得图 3.18。从图 3.18 中可以看出，教育与就业、社会与医疗保险相关的帖子所占比例在下半年有所上升，收入分配相关的帖子所占比例在下半年有所下降；同时与两个主题相关的帖子所占比例没有明显变化。这个结果和作者所属主题的比例图相一致。

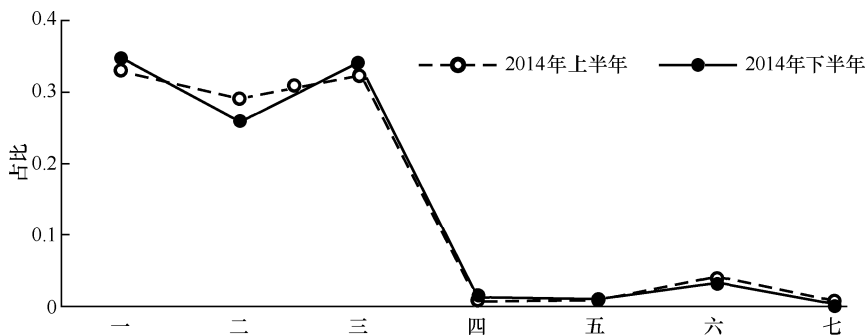


图 3.18 各主题的帖子占比

一、教育与就业讨论参考者；二、收入分配讨论参考者；三、社会与医疗保障讨论参考者；  
四、教育与就业、收入分配讨论参考者；五、教育与就业、社会与医疗讨论参考者；  
六、收入分配、社会与医疗保障讨论参考者；七、教育与就业、收入分配、社会与医疗保障讨论参考者

基于客体对象的社会网络基本指标如表 3.9 所示。2014 年上半年和 2014 年下半年的指标基本相似。值得注意的是，这图 3.16 和图 3.17 的模块化指标都比较高，说明了这两张图的社区划分质量都比较高，这是由于这三个主题的相关性较弱，属于两个或三个主题的帖子所占比例相对较低，对社区划分的影响较小。

表 3.9 网络基本指标对比

调研时间	平均度	网络直径	图密度	模块化	平均路径长度
2014 年上半年	2.107	4	0.001	0.615	3.260
2014 年下半年	2.095	4	0.001	0.619	3.261

3.5 生态社区网络特征分析

当今社会，伴随着经济的飞速发展，生态问题日益明显，对生态领域下的舆情问题的探索显得越来越有意义。本节对生态领域整体数据进行了介绍，并且对该领域的上半年和下半年所有主体和客体在三大主题下的分布情况进行了可视化，并从中发现量的演化规律。

3.5.1 生态领域数据视图

该研究领域的数据中，帖子总数为 36807 条，评论总数为 501239 条，参与发帖的总人数为 17954 人，参与评论的总人数为 107587 人，具体的分布如表 3.10 所示。

表 3.10 生态领域数据

主题分类	发帖数/条	评论数/条	发帖人数/个	评论人数/个
生态环境	10074	101607	6168	32022
生态资源	15341	309333	7923	64852
技术或措施	11392	90299	7023	24279
合计	36807	501239	17954	107587

3.5.2 基于主体对象的网络演化分析

由节 3.1.1 中提出的建模过程构建生态领域基于主体对象的社会网络分别如图 3.19 和图 3.20 所示。

图 3.19 是根据 2014 年上半年数据构建论坛中较活跃的用户参与生态领域中热点主题的社会网络图。可以发现，讨论生态资源的活跃用户稍多于讨论生态环境及讨论技术措施的用户，讨论生态环境的用户与讨论技术措施的用户数量相差不大，只有少数用户同时活跃在生态领域的三大主题下。我们可以将用户分为三大类：第一类用户参与三个主题的讨论；第二类用户参与两个主题的讨论；第三类用户仅仅参与其中一个主题的讨论。

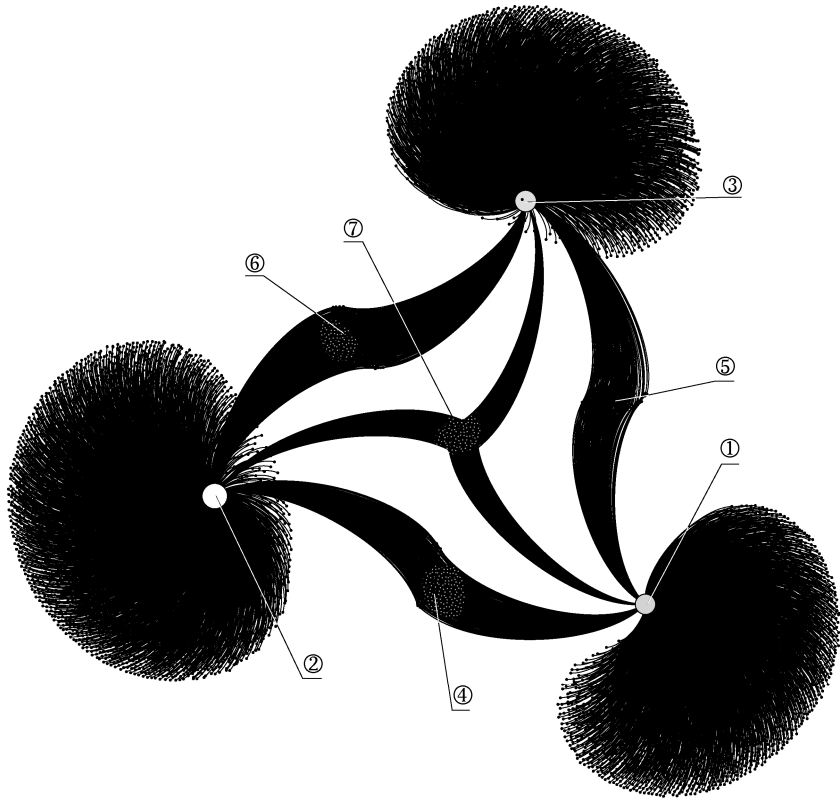


图 3.19 (2014 年上半年) 基于人的主题网络 (2014.1.1~2014.6.30)

- ① 仅讨论生态环境的作者数量为 2461 人, 占比为 0.255;
- ② 仅讨论生态资源的作者数量为 3325 人, 占比为 0.345;
- ③ 仅讨论技术与措施的作者数量为 2638 人, 占比为 0.273;
- ④ 同时讨论生态环境、生态资源的作者数量为 312 人, 占比为 0.032;
- ⑤ 同时讨论生态环境、技术与措施的作者数量为 288 人, 占比为 0.030;
- ⑥ 同时讨论生态资源、技术与措施的作者数量为 374 人, 占比为 0.039;
- ⑦ 同时讨论生态环境、生态资源、技术与措施的作者数量为 248 人, 占比为 0.026

第一类用户的人数最少, 他们参与了生态领域所有热点主题的讨论, 这部分人他们是生态领域的中流砥柱, 最有可能成为生态领域的意见领袖。第二类用户是生态领域的重要参与者, 他们经常在网上参与某些热点话题的讨论。第三类用户可能单纯对某一热点话题感兴趣, 这种用户最有可能发生变动。

图 3.20 是根据 2014 年下半年数据构建论坛中较活跃的作者参与生态领域中热点主题的社会网络图。可以发现, 生态领域的参与作者总体数量变化不大,

各主题所占比列较为稳定。活跃在生态环境和生态资源的作者都有所减少，在技术措施领域的活跃作者有所增加，可能是以前的关注生态环境和生态资源的部分作者开始关注改变环境及节约资源的措施或技术。

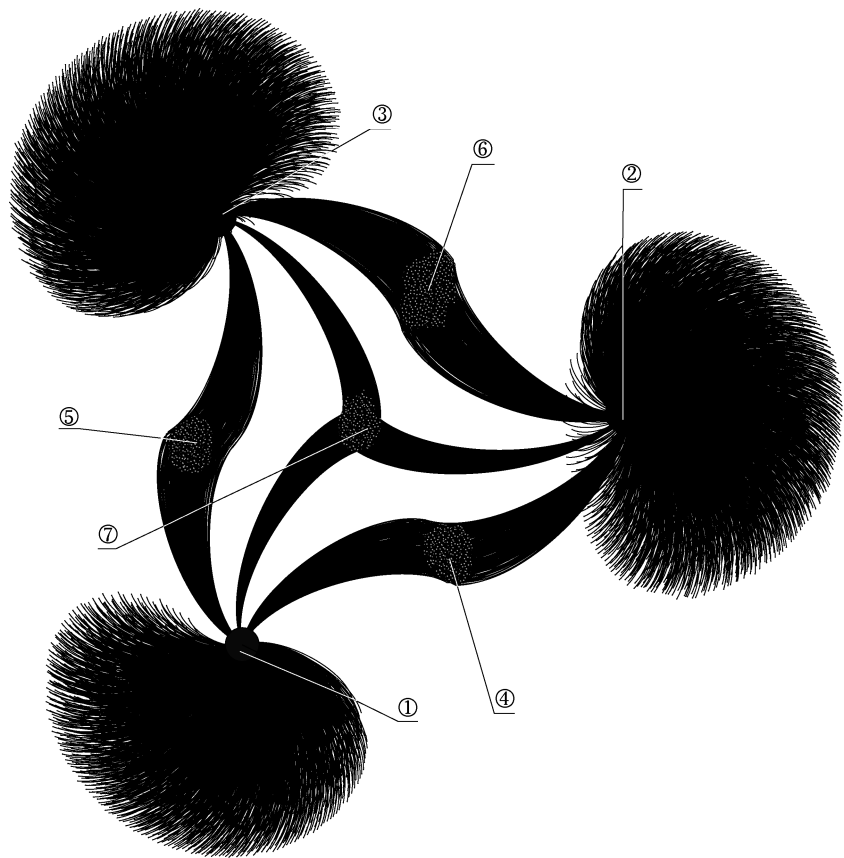


图 3.20 （2014 年下半年）基于人的主题网络（2014.7.1~2014.12.31）

- ① 仅讨论生态环境的作者数量为 2207 人，占比为 0.235；
- ② 仅讨论生态资源的作者数量为 3145 人，占比为 0.334；
- ③ 仅讨论技术与措施的作者数量为 2722 人，占比为 0.289；
- ④ 同时讨论生态环境、生态资源的作者数量为 276 人，占比为 0.029；
- ⑤ 同时讨论生态环境、技术与措施的作者数量为 311 人，占比为 0.033；
- ⑥ 同时讨论生态资源、技术与措施的作者数量为 409 人，占比为 0.044；
- ⑦ 同时讨论生态环境、生态资源、技术与措施的作者数量为 336 人，占比为 0.036

将图 3.19 与图 3.20 中参与生态领域讨论帖子的占比绘图得图 3.21。从图 3.21 中可以看出，关注技术措施的作者所占比例在下半年有明显上升，关

注生态环境和生态资源的作者所占比例在下半年有明显下降。同时可以看出，大部分作者仅关注一个主题，说明广大网民的关注点较为单一；关注两个甚至三个主题的作者很有可能成为生态领域的意见领袖，他们的言论将对其他网民产生较大的影响。

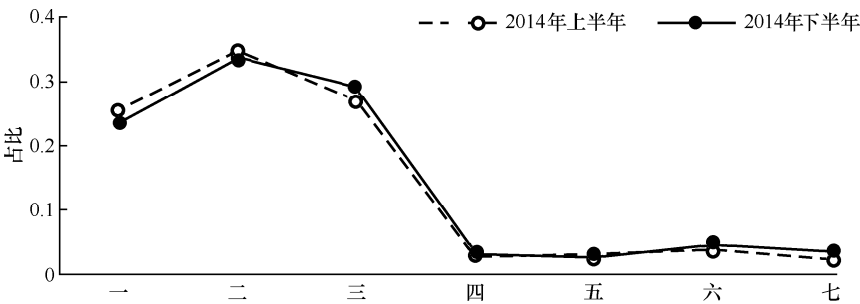


图 3.21 参与各主题的用户占比

一、生态环境讨论参考者；二、生态资源讨论参考者；三、技术或措施讨论参考者；  
四、生态环境、生态资源讨论参考者；五、生态环境、技术或措施讨论参考者；  
六、生态资源、技术或措施讨论参考者；七、生态环境、生态资源、技术或措施讨论参考者

基于主体对象的社会网络基本指标如表 3.11 所示。2014 年上半年和 2014 年下半年的指标基本相似。平均度下半年略大于上半年，说明下半年的网民“综合度”更大，即下半年网民的帖子平均涉及的主题更多；网络直径都为 4，说明任意两个节点之间最大的距离为 4；图密度都为 0，说明图 3.19 和图 3.20 都是稀疏图；模块化指标分别为 0.512 和 0.517，说明图 3.20 划分的社区质量更高，这也说明下半年同属于两个或三个以上社区的作者相对较少；平均路径长度分别为 3.126 和 3.104，基本相似。

表 3.11 网络基本指标对比

调研时间	平均度	网络直径	图密度	模块化	平均路径长度
2014 年上半年	2.301	4	0.000	0.512	3.126
2014 年下半年	2.338	4	0.000	0.517	3.104

3.5.3 基于客体对象的网络演化分析

由节 3.1.2 中提出的建模过程构建生态领域基于客体对象的社会网络分别如图 3.22 和图 3.23 所示。



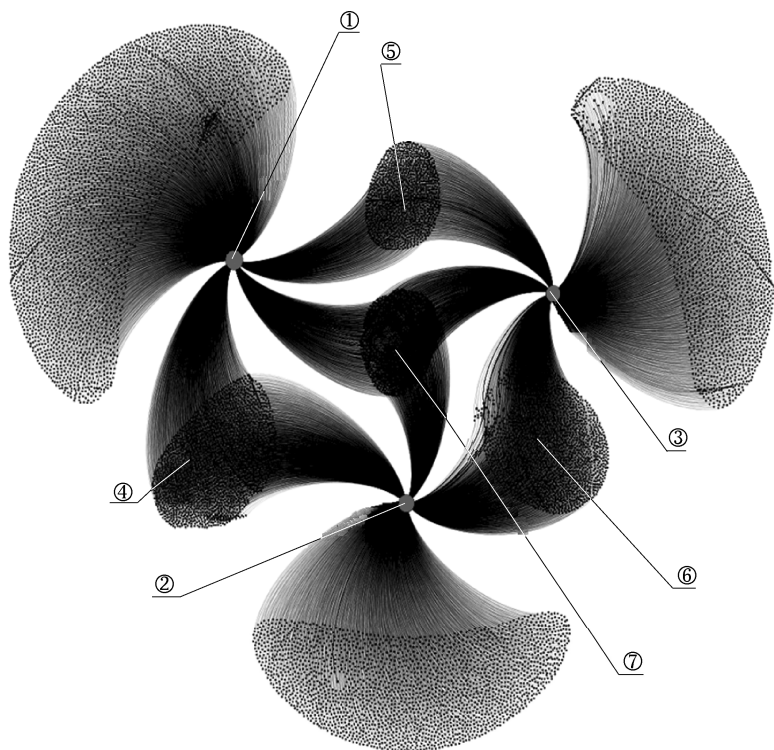


图 3.22 (2014 年上半年) 基于内容的主题网络 (2014.1.1~2014.6.30)

- ① 仅讨论生态环境的帖子数量为 2980 个, 占比为 0.168;
- ② 仅讨论生态资源的帖子数量为 4752 个, 占比为 0.268;
- ③ 仅讨论技术与措施的帖子数量为 2775 个, 占比为 0.157;
- ④ 同时讨论生态环境、生态资源的帖子数量为 1166 个, 占比为 0.066;
- ⑤ 同时讨论生态环境、技术与措施的帖子数量为 2207 个, 占比为 0.125;
- ⑥ 同时讨论生态资源、技术与措施的帖子数量为 2165 个, 占比为 0.122;
- ⑦ 同时讨论生态环境、生态资源、技术与措施的帖子数量为 1669 个, 占比为 0.094

图 3.22 是根据 2014 年上半年数据构建生态领域中热点主题的社会网络图。可以发现, 单独讨论生态环境和生态措施帖子数量大致相当, 生态资源的帖子数量略大于其他两个主题, 说明网民对于生态资源的关注度比其他两个领域高。

同时可以发现, 论坛中帖子具有很强的综合倾向, 即一个帖子讨论的内容涉及两个甚至三个主题。同时涉及生态环境和技术与措施两个主题的帖子数量与同时涉及生态环境和技术与措施两个主题的帖子数量都比较多, 反映出技术与措施和生态环境、生态资源都具有较强的关联性; 同时涉及三个主题的帖子数量较多, 反映出部分网民对于生态领域的综合了解, 这些网民可能会成为生态领域的意见领袖。

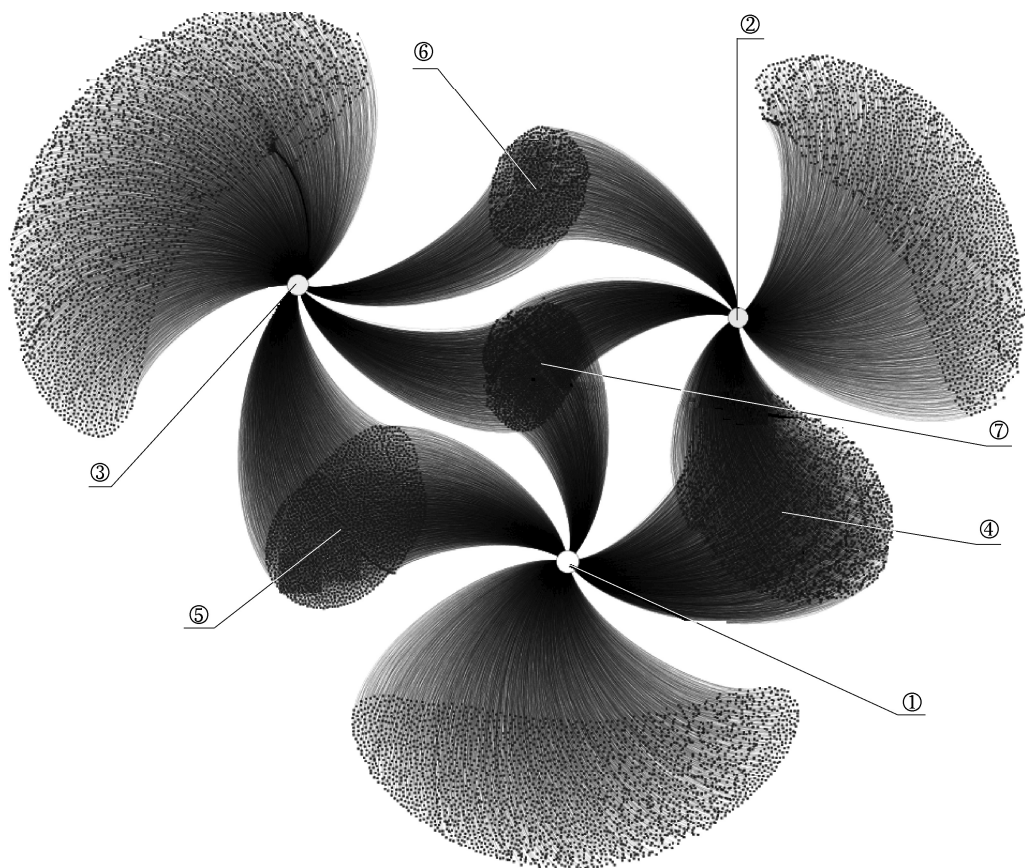


图 3.23 (2014 年下半年) 基于内容的主题网络 (2014.7.1~2014.12.31)

- ① 仅讨论生态环境的帖子数量为 3073 个, 占比为 0.162;
- ② 仅讨论生态资源的帖子数量为 4545 个, 占比为 0.239;
- ③ 仅讨论技术与措施的帖子数量为 3299 个, 占比为 0.174;
- ④ 同时讨论生态环境、生态资源的帖子数量为 1104 个, 占比为 0.058;
- ⑤ 同时讨论生态环境、技术与措施的帖子数量为 2954 个, 占比为 0.155;
- ⑥ 同时讨论生态资源、技术与措施的帖子数量为 2365 个, 占比为 0.125;
- ⑦ 同时讨论生态环境、生态资源、技术与措施的帖子数量为 1647 个, 占比为 0.087

图 3.23 是根据 2014 年下半年数据构建生态领域中热点主题的社会网络图。可以发现, 讨论生态资源与生态环境两个主题的帖子数量与上半年基本持平, 讨论技术与措施的帖子数量有明显增多, 说明大多数帖子已经开始关注解决生态问题的技术或措施; 帖子所属的主题分布也与上半年相似, 但涉及生态环境和技术与措施两个主题的帖子数量明显增多, 说明对于环境问题的解决措施的帖子明显增多。

将图 3.22 与图 3.23 中参与生态领域讨论帖子的占比绘图得图 3.24。从图 3.24 中可以看出，生态领域三大主题的相关帖子数量在下半年大部分有所上升；其中与技术措施相关的帖子数量明显增多，比例也明显增加，说明网民对于环境问题的解决办法关注度有所增强；而与生态资源主题相关的帖子数量明显减少，比例也明显减少，说明下半年网民对于生态资源关注度有所下降；生态环境主题相关帖子的增加也说明了环境问题正在日益凸显；生态环境与技术措施两个主题相关的帖子数量及比例明显增加，说明了网民对于环境问题的关注以及对环境问题的解决办法的关注有明显增加。

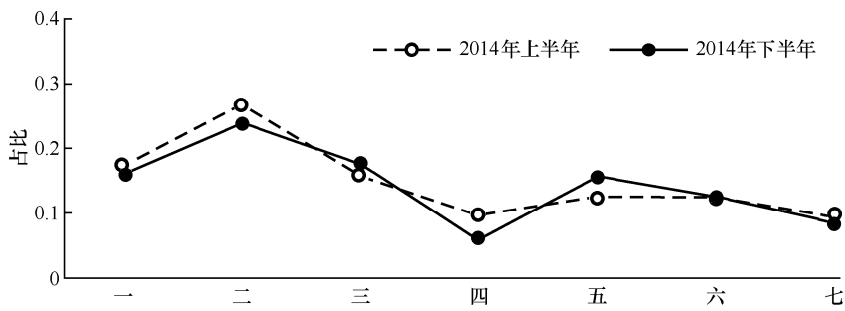


图 3.24 各主题的帖子占比

一、生态环境讨论参考者；二、生态资源讨论参考者；三、技术或措施讨论参考者；  
四、生态环境、生态资源讨论参考者；五、生态环境、技术或措施讨论参考者；  
六、生态资源、技术或措施讨论参考者；七、生态环境、生态资源、技术或措施讨论参考者

基于客体对象的社会网络基本指标如表 3.12 所示。2014 年上半年和 2014 年下半年的指标基本相似。平均度下半年略大于上半年，说明下半年的帖子“综合度”更大一点，即下半年帖子平均描述的主题更多；网络直径都为 4，说明任意两个节点间最大的距离为 4；图密度都为 0 说明图 3.4 和图 3.5 都是稀疏图；模块化指标分别为 0.331 和 0.327，总体上是较低的指标值，说明这两个网络中的社区不明显，不易被发现；平均路径长度为 2.709 和 2.703，说明帖子数相差不大。

表 3.12 网络基本指标对比

调研时间	平均度	网络直径	图密度	模块化	平均路径长度
2014 年上半年	3.009	4	0.000	0.331	2.709
2014 年下半年	3.028	4	0.000	0.327	2.703

### 3.6 国际和军事社区网络特征分析

随着互联网的发展，信息的流通越来越迅速，网民在网上冲浪时，会遇到各种各样的国际新闻，或许是灾难性的，或许是合作性的，或者是国际上国家为了维护自己国家的主权和领土的完整而发动战争的，这些问题都说明虽然在这个看似和平的年代，各种大事件都在发生。因此，本节对国际和军事领域整体数据进行了介绍，并且对该领域所有主体和客体在三大主题下的分布情况进行了可视化，从中发现量上的演化规律。

#### 3.6.1 国际和军事领域数据视图

本研究领域的数据中，帖子总数为 6964 条，评论总数为 246723 条，参与发帖的总人数为 5967 人，参与评论的总人数为 33035 人，具体的分布如表 3.13 所示。

表 3.13 国际和军事领域数据视图

主题分类	发帖数/条	评论数/条	发帖人数/个	评论人数/个
合作共荣	790	30403	642	4352
国际事故	1145	44488	1004	6796
领土主权	2009	91998	1849	9342
其他	3020	79843	2472	12545
合计	6964	246723	5967	33035

#### 3.6.2 基于主体对象的网络演化分析

由节 3.1.1 中提出的建模过程构建国际和军事领域基于主体对象的社会网络分别如图 3.25 和图 3.26 所示。

图 3.25 是根据 2014 年上半年数据构建论坛中较活跃的用户参与国际和军事领域中热点主题的社会网络图。可以发现，讨论领土主权、国际事故两个主题的活跃用户数量多于讨论合作共荣主题的用户数量，只有少数用户同时活跃在国际与军事领域的三大主题下。我们可以将用户分为三大类：第一类用户同时参与三个主题的讨论；第二类用户同时参与两个主题的讨论；第三类用户仅参与其中一个主题的讨论。

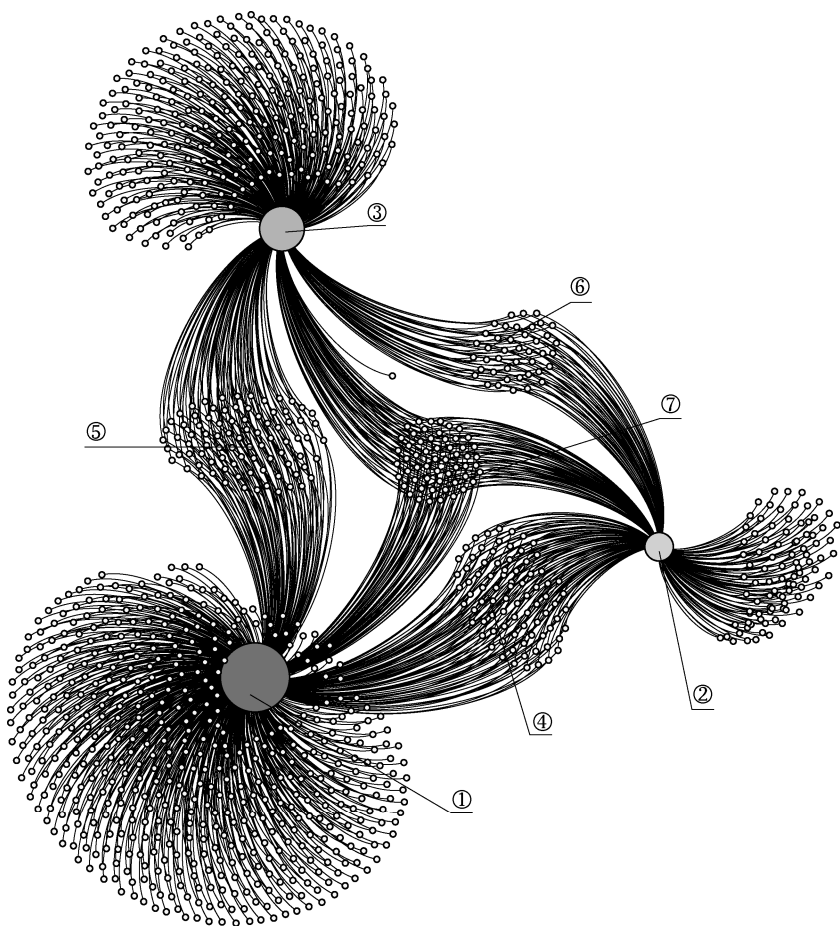


图 3.25 (2014 年上半年) 基于人的主题网络 (2014.1.1~2014.6.30)

- ① 仅讨论领土主权的作者数量为 681 人, 占比为 0.480;
- ② 仅讨论合作共赢的作者数量为 187 人, 占比为 0.132;
- ③ 仅讨论国际事故的作者数量为 354 人, 占比为 0.250;
- ④ 同时讨论领土主权、合作共赢的作者数量为 44 人, 占比为 0.031;
- ⑤ 同时讨论领土主权、国际事故的作者数量为 78 人, 占比为 0.055;
- ⑥ 同时讨论合作共赢、国际事故的作者数量为 65 人, 占比为 0.046;
- ⑦ 同时讨论领土主权、合作共赢、国际事故的作者数量为 9 人, 占比为 0.006

第一类用户的人数最少, 他们参与了国际和军事领域所有热点主题的讨论, 这部分人是国际和军事领域的中流砥柱, 最有可能成为国际和军事领域的意见领袖。第二类用户是国际与军事领域的重要参与者, 他们经常在网上参与某些热点话题的讨论。第三类用户可能单纯对某一热点话题感兴趣, 这种用户最有可能发生变动, 因为他们参与某一话题最大的因素是阶段性的需求。但是,

也有一部分人可能是国际与军事领域某一主题的长期参与者，如有很强的国家荣辱观的用户可能关注的永远只是领土主权的相关信息。

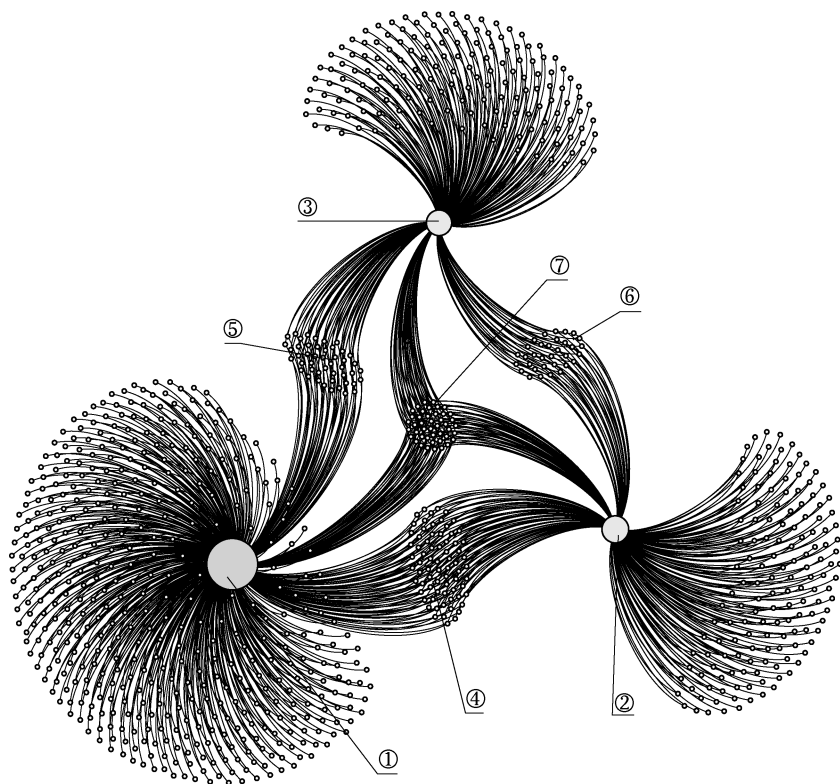


图 3.26 (2014 年下半年) 基于人的主题网络 (2014.7.1~2014.12.31)

- ① 仅讨论领土主权的作者数量为 594 人，占比为 0.508；
- ② 仅讨论合作共赢的作者数量为 256 人，占比为 0.219；
- ③ 仅讨论国际事故的作者数量为 240 人，占比为 0.205；
- ④ 同时讨论领土主权、合作共赢的作者数量为 25 人，占比为 0.021；
- ⑤ 同时讨论领土主权、国际事故的作者数量为 18 人，占比为 0.015；
- ⑥ 同时讨论合作共赢、国际事故的作者数量为 31 人，占比为 0.026；
- ⑦ 同时讨论领土主权、合作共赢、国际事故的作者数量为 6 人，占比为 0.006

图 3.26 是根据 2014 年下半年数据构建论坛中较活跃的用户参与国际和军事领域中热点主题的社会网络图。可以发现，国际和军事领域的参与用户总体数量维持基本不变，各主题所占比例较为稳定，只有国际事故领域下的帖子数量变少了。国际事故主题所占比例的下降的原因可能是上半年马航 MH370 失联引起了广大网友的关注，并且在论坛上引起了激烈的讨论和对马航失联事件的各种猜测。

将图 3.25 与图 3.26 中参与国际和军事领域讨论帖子的占比绘图得图 3.27。从图 3.27 中可以看出，关注国际事故主题下的作者所占比例在下半年有明显下降，关注合作共荣主题下的作者所占比例在下半年有明显上升。同时可以看出，大部分作者仅关注一个主题，说明广大网民的关注点较为单一；关注两个甚至三个主题的作者很有可能成为国际与军事领域的意见领袖，他们的言论将对其他网民产生较大的影响。

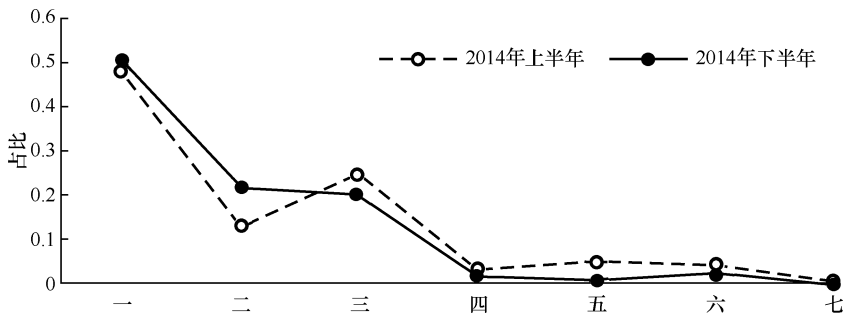


图 3.27 参与各主题的用户占比

一、领土主权讨论参考者；二、合作共荣讨论参考者；三、国际事故讨论参考者；  
四、领土主权讨论参考者、合作共荣讨论参考者；五、领土主权、国际事故讨论参考者；  
六、合作共荣、国际事故讨论参考者；七、领土主权、合作共荣、国际事故讨论参考者

基于主体对象的社会网络基本指标如表 3.14 所示。2014 年上半年和 2014 年下半年的指标基本相似。平均度上半年略大于下半年，说明上半年的网民“综合度”更大，即上半年网民的帖子平均涉及的主题更多；网络直径都为 4，说明任意两个节点之间最大的距离为 4；图密度都比较小，说明图 3.25 和图 3.26 都是稀疏图；模块化指标分别为 0.430 和 0.446，说明图 3.26 划分的社区质量更高，这也说明下半年同属于两个或三个以上社区的作者相对较少；平均路径长度分别为 2.795 和 2.852，下半年的平均长度较大说明下半年的帖子数相对有所增加。

表 3.14 网络基本指标对比

调研时间	平均度	网络直径	图密度	模块化	平均路径长度
2014 年上半年	2.560	4	0.002	0.430	2.795
2014 年下半年	2.493	4	0.003	0.446	2.852

3.6.3 基于客体对象的网络演化分析

由节 3.1.2 中提出的建模过程构建国际与军事领域基于客体对象的社会网络分别如图 3.28 和图 3.29 所示。

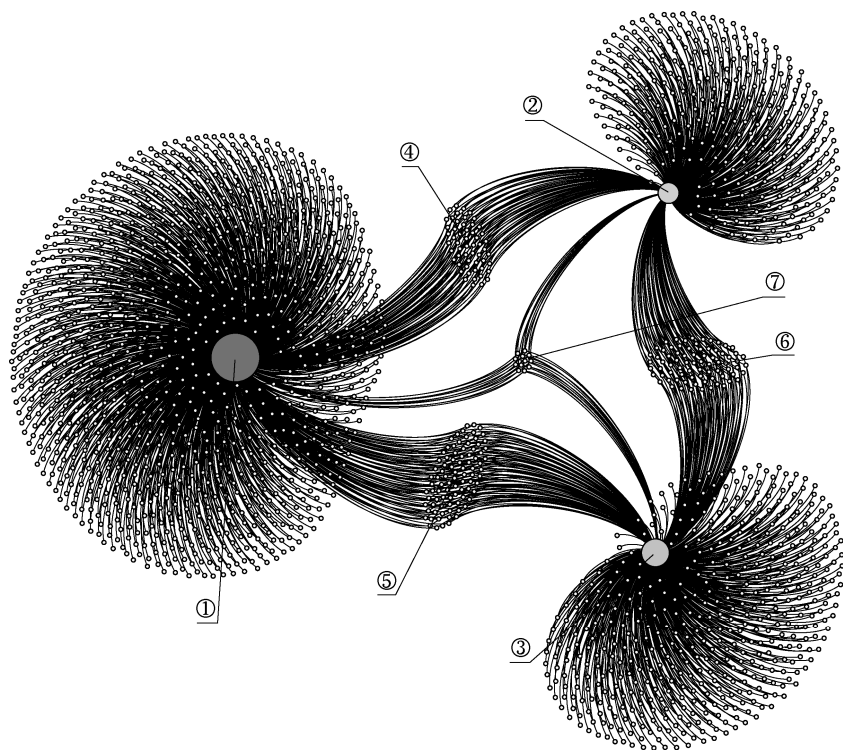


图 3.28 (2014 年上半年) 基于内容的主题网络 (2014.1.1~2014.6.30)

- ① 仅讨论领土主权的帖子数量为 1068 个, 占比为 0.510;
- ② 仅讨论合作共赢的帖子数量为 318 个, 占比为 0.152;
- ③ 仅讨论国际事故的帖子数量为 476 个, 占比为 0.227;
- ④ 同时讨论领土主权、合作共赢的帖子数量为 55 个, 占比为 0.026;
- ⑤ 同时讨论领土主权、国际事故的帖子数量为 88 个, 占比为 0.043;
- ⑥ 同时讨论合作共赢、国际事故的帖子数量为 76 个, 占比为 0.036;
- ⑦ 同时讨论领土主权、合作共赢、国际事故的帖子数量为 12 个, 占比为 0.006

图 3.28 是根据 2014 年上半年数据构建国际和军事领域中热点主题的社会网络图。可以发现, 单独讨论领土主权、国际事故、合作共赢三个主题的帖子数量大致相当, 领土主权的数量略大于其他两个主题, 说明我国网民具有强烈的民族责任感。同时可以发现, 论坛中的帖子具有很强的综合倾向, 即一个帖子讨论的内容涉及两个甚至三个主题。同时涉及国际事故和领土主权两个主题的帖子数量相对较多, 反映出这两个主题之间具有较强的关联性; 同时涉及领土主权与合作共赢、国际事故与合作共赢的帖子数量相对来说较少, 显而易见, 这两个领域联系较少, 交集不多。而同时关注三个专题的帖子数更少, 说明帖子同时涉及这三个领域的数量较少, 反映出部分帖子



发帖的网民对于国际和军事领域的综合了解，这些网民可能会成为国际和军事领域的意见领袖。

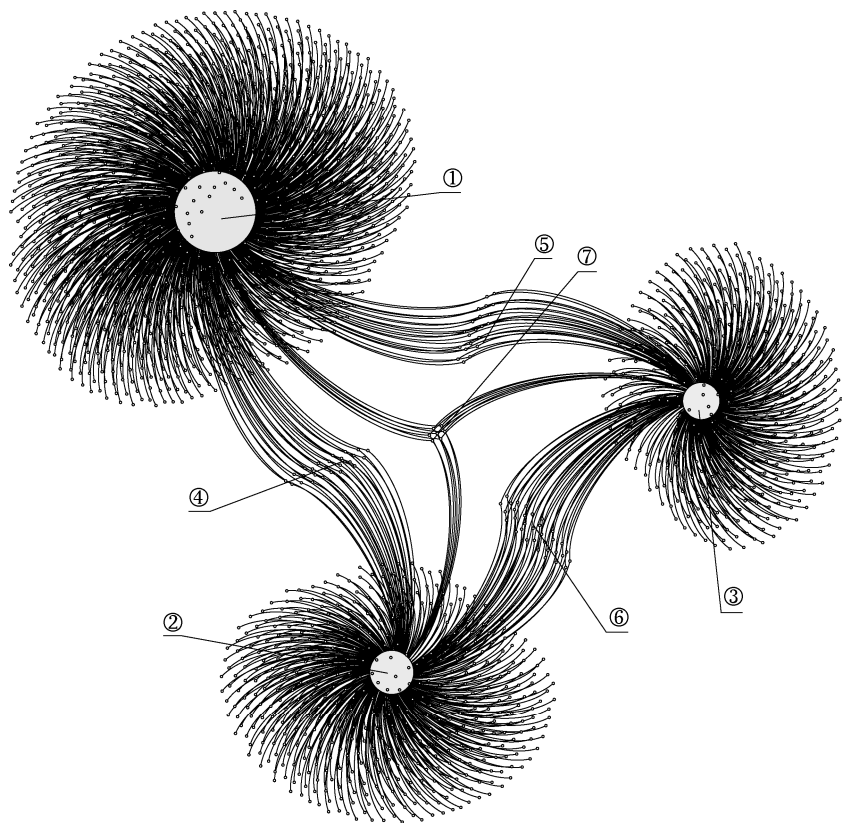


图 3.29 (2014 年下半年) 基于内容的主题网络 (2014.7.1~2014.12.31)

- ① 仅讨论领土主权的帖子数量为 927 个，占比为 0.505；
- ② 仅讨论合作共赢的帖子数量为 451 个，占比为 0.246；
- ③ 仅讨论国际事故的帖子数量为 370 个，占比为 0.202；
- ④ 同时讨论领土主权、合作共赢的帖子数量为 25 个，占比为 0.014；
- ⑤ 同时讨论领土主权、国际事故的帖子数量为 21 个，占比为 0.011；
- ⑥ 同时讨论合作共赢、国际事故的帖子数量为 33 个，占比为 0.018；
- ⑦ 同时讨论领土主权、合作共赢、国际事故的帖子数量为 7 个，占比为 0.004

图 3.29 是根据 2014 年下半年数据构建国际和军事领域中热点主题的社会网络图。可以发现，讨论领土主权、合作共赢、国际事故三个主题的帖子数量与上半年基本持平，帖子所属的主题分布也与上半年相似，即同时涉及领土主权和国际事故两个主题的帖子数量较多；同时涉及这三个主题的帖子数量较少。

将图 3.28 与图 3.29 中参与国际和军事领域讨论帖子的占比绘图得图 3.30。从图 3.30 中可以看出,关注合作共赢的作者所占比例在下半年有明显上升,关注另外两个专题的作者所占比例在下半年有略有下降,同时参与两个主题讨论与同时参与三个主题的作者比例变化略有下降,同时参与三个主题的作者比例与上半年相比基本持平。

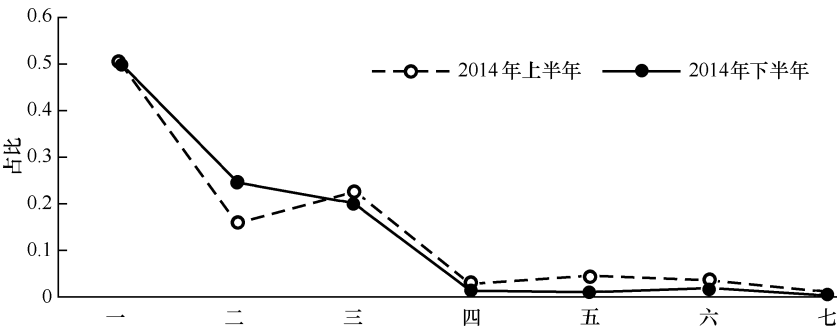


图 3.30 各主题的帖子占比

一、领土主权讨论参考者；二、合作共赢讨论参考者；三、国际事故讨论参考者；  
四、领土主权讨论参考者、合作共赢讨论参考者；五、领土主权、国际事故讨论参考者；  
六、合作共赢、国际事故讨论参考者；七、领土主权、合作共赢、国际事故讨论参考者

基于客体对象的社会网络基本指标如表 3.15 所示。2014 年上半年和 2014 年下半年的指标基本相似。平均度上半年略大于下半年,说明上半年的帖子“综合度”更大一点,即上半年帖子平均描述的主题更多;网络直径都为 4,说明任意两个节点间最大的距离为 4;图密度都为 0.001 说明图 3.28 和图 3.29 都是稀疏图;模块化指标分别为 0.524 和 0.579,上半年的指标值说明这两个网络中的社区不明显,不易被发现,下半年社区划分趋势更明显;平均路径长度为 3.034 和 3.16,下半年的平均长度较大说明下半年的帖子数增多。

表 3.15 网络基本指标对比

调研时间	平均度	网络直径	图密度	模块化	平均路径长度
2014 年上半年	2.229	4	0.001	0.524	3.034
2014 年下半年	2.098	4	0.001	0.579	3.160

3.7 本章小结

本章对整个天涯论坛的原始帖子数据进行了分类,在每个领域下再分别选

择几大热点主题，构建特征向量，对该领域内的数据再次进行分类，根据分类结构构建领域内的基于主体对象和客体对象的社会网络，从而发现各个领域内用户的讨论热点。本章从内容分布统计和社会网络分析两个角度对各个领域的整体数据进行分析，使读者对各个领域的热点主题的帖子及参与作者等情况有一个整体的了解。

## 第 4 章 经济领域的相关舆情分析

1978 年，发端于中国共产党十一届三中全会的改革开放，改变了中国，影响了世界。此后 35 年间的 7 次三中全会，改革是贯穿其中的不变的主题。也是因为如此，在中国的政治词汇中，“三中全会”几乎成为改革开放的代名词。在新的历史条件下，现实的发展要求我们转换经济体制改革的视角，需要形成新思路。

以经济建设为中心，是党的基本路线方针，是我国经济社会发展的一条主线。我们要坚持和完善基本经济制度，加快完善现代市场体系，宏观调控体系，开放型经济体系，加快转变经济发展方式，推动经济更有效率，更加公平，更可持续发展。

为了客观地分析经济领域的热点舆情，我们从天涯论坛中选取了 2014 年度一些经济领域热点专题的帖子进行深入的总结和分析，本章后续部分将从不同的角度展示该领域舆情分析的结果。

### 4.1 经济领域热点专题和热点词分析

本节选取经济领域的相关帖子进行分析，通过对词频进行统计得到该领域的热点词及演化情况，通过对原帖进行文本聚类得到热点事件，从而确定该领域的热点专题。

#### 4.1.1 经济领域热点词分析

热点词出现的频率会随着热点事件的发生而快速增长，从热点词的演化过程可以看出不同时间热点事件的主题变化情况。对 2014 年全年的数据按季度进行词频统计和排名，结果如表 4.1 所示。

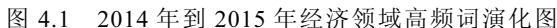
表 4.1 高频词汇统计

2014.01~2014.03		2014.04~2014.06		2014.07~2014.09		2014.10~2014.12		2015.01~2015.03	
词语	词频	词语	词频	词语	词频	词语	词频	词语	词频
机器人	248	中国	347	中国	257	中国	536	中国	10935
中国	215	经济	242	经济	232	美国	471	经济	10570
经济	207	美国	185	美国	180	经济	467	美国	8293

续表

2014.01~2014.03		2014.04~2014.06		2014.07~2014.09		2014.10~2014.12		2015.01~2015.03	
词语	词频	词语	词频	词语	词频	词语	词频	词语	词频
美国	154	人口	155	银行	162	房地产	265	人民币	6700
房地产	126	房地产	143	资本	150	人民币	266	房地产	6692
人民币	120	人民币	135	人民币	147	资本	265	资本	6658
汇率	98	股市	116	房地产	146	石油	208	汇率	5145
加息	98	政策	113	汇率	112	银行	205	加息	5141
电机	98	银行	108	加息	112	加息	202	银行	4929
行业	96	数据	103	阿里	103	市场	197	出台	4410
编码器	94	汇率	99	金融	91	投资	193	投资	3933
人民币	84	加息	98	投资	90	汇率	190	收益	3799
银行	84	股票	89	市场	87	美元	147	政策	3743
政策	71	股市	88	收益	86	上涨	145	外流	3660
大陆	70	社会	83	政策	85	指数	145	时间	3506
投资	70	报道	82	大陆	81	股票	143	市场	3478
外流	70	生育	74	外流	81	政策	143	股市	3310
收益	70	收益	73	涨	77	大陆	142	股票	3300
市场	64	大陆	72	出台	76	外流	138	美元	3002
项目	61	市场	72	股市	73	理论	137	住房	2956
美元	56	外流	70	降	72	涨	135	改善	2951
住房	56	放开	69	部长	65	主席	114	FOMC	2929
股市	56	生育率	67	美元	64	公司	113	指数	2390
FOMC	56	房价	65	住房	64	规则	113	回调	2383
改善	56	美元	60	FOMC	64	行业	112	政府	2380
主席	56	住房	57	主席	64	黄金	111	金融	2323
官方	56	规则	57	行业	58	住房	109	房子	2319
社会	56	部长	56	房子	55	FOMC	108	指标	2308
政府	47	FOMC	56	压力	54	政府	105	GDP	2305
预期	44	主席	56	股票	54	发生	103	行业	2294
饭碗	44	指标	54	放开	54	房价	103	债	2236
群体	43	孩子	53	政府	52	金融	98	美联储	2235
股票	43	政府	52	花粉	50	上调	92	QE	2215
GDP	42	调查	50	回调	50	下跌	90	调查	2214

为了更好地观察高频词的演化过程,我们用作图工具 NetDraw 对高频词和时间的对应关系进行可视化,得到了高频词演化图(图 4.1)。



#### 4.1.2 经济领域热点专题分析

从表 4.2 可以看出, 经济领域的热点事件中股票投资、经济危机、房地产和汇率等话题关注度较高。这是因为房地产、股票、经济危机和汇率等经济问

题和百姓生活状况息息相关。本报告以房市与房价、股市与股票、货币与汇率三个专题为例，对经济领域的网络舆情进行深入分析。

表 4.2 经济领域相关的十大热点子专题

热点子专题的名称	帖子量/条	总浏览量/条	总回复量/条	参与人数/个	专题热度
俄罗斯是中美经济的平衡点	4	1189006	9680	4756	0.999998
每日荐一牛股，送于有缘人	3	81017	2480	777	0.794319
一个笔记本电脑小卖家与大家聊聊二手笔记本的出路	3	74412	1628	643	0.777814
风雨飘摇的零售行业	1	275108	1150	785	0.739344
南阳一个丘陵地区建设	1	140722	958	624	0.666732
经济与股票投资	4	102306	690	451	0.653868
石油券商银行将全面暴跌	4	145632	654	386	0.596215
经济危机的乌云正悄然袭来	3	283271	590	427	0.579346
美国资本市场那些事	5	23661	538	412	0.549970
人民币与美元汇率	4	13608	170	96	0.483275

4.2 经济领域意见领袖分析

在经济领域中涉及的用户包括相关帖子的发布者以及参与浏览和回复这些帖子的舆情受众。在信息传播中，一些网民凭借自身表达能力突出、见解深刻等素质，成为政治领域中的“意见领袖”。他们善于从表象信息透析到比较深层的价值内涵和意义，并能对问题进行解析，形成自己的观点并加以表达，他们的观点能够引起众多普通网民的兴趣和参与，促成网络舆论的最终形成。

通过对网络舆情发布者的关注度分析，可以有效地监测出网络舆情在传播与演化过程中具有影响力的意见领袖。本节采用发布者影响力、发布者活跃度和发布者价值这三个不同维度的指标对舆情发布者进行分析，采用负面指数、参与度和受众影响力对各个专题的舆情受众进行分析。

4.2.1 意见领袖

1) 发布者影响力

计算公式见节 2.3.2。计算出的影响力结果如表 4.3 所示。

表 4.3 经济领域论坛用户影响力排名

用户	粉丝数/个	发帖数/条	帖子点击数/次	帖子回复数/条	回帖数/条	影响力
骑白马的农民	7821	1	9049518	58634	16	0.995054
chenkx214	1172	2	2239062	18127	624	0.634769
贫道贫得很	5393	1	1638993	5456	38	0.503493
生于 0715	28196	2	1318363	2642	262	0.478845
大江宁静	2063	8	1508915	1564	1493	0.478445
258308910	779	1	1400018	541	27	0.432635
三高的浪人	5944	1	1335958	264	24	0.426541
stupidwz	1651	1	1189006	311	1140	0.396616
c1698477744	1030	1	1256039	154	5	0.393889
泡泡破掉了	1565	1	849008	123	383	0.282563
老农今年又白忙 1	3437	4	698914	74	1755	0.247601
黄汉升 1	752	1	652355	56	1252	0.231084
天涯财经	74904	2	88354	51	7	0.223670
williamm916	19	6	644613	24	507	0.218145
陕北黑脸汉子	48310	4	219568	18	500	0.198896

表 4.3 列出了用户影响力前 15 位的用户。“骑白马的农民”的影响力最高，同时回复数和浏览次数都较高，可将其视作意见领袖；“老农今年又白忙”和“陕北黑脸汉子”在之前的发布者影响力排名中也出现过。从表 4.3 以看出：浏览次数由于数量较大，对于用户影响力也大，此种情况不合理，公式还需要进一步改进。

## 2) 发布者活跃度

公式具体解释见节 2.3.2。这里， $g_1$  取 0.6， $g_2$  取 0.4，计算结果如表 4.4 所示。

表 4.4 经济领域论坛用户活跃度排名

用户	发帖数/条	回帖数/条	活跃度
关羽 a	2	2717	1.000000000
坚守信念钦 2014	1	2164	0.796139706
本座老姐拜斑竹	1	1906	0.701286765
庄梓 2014	1	1869	0.687683824
火哥闯天下	1	1774	0.652757353



续表

用户	发帖数/条	回帖数/条	活跃度
老农今年又白忙 1	4	1755	0.647426471
大江宁静	8	1493	0.553308824
先人 jzy	0	1481	0.544485294
万古枯尘	1	1431	0.526654412
小媛媛 A	1	1371	0.504595588
老秦人 1990	1	1337	0.492095588
秋风飞扬 123	1	1288	0.474080882
空忍者	2	1281	0.472058824
黄汉升 1	1	1252	0.460845588
狸猫 2301	0	1242	0.456617647

表 4.4 列出了发布者活跃度前 15 名的用户，表明这些用户在论坛中较为活跃，可以作为重点监视的对象。“老农今年又白忙 1”和“大江宁静”属于活跃性的意见领袖，他们保持着较高的发帖量和回帖量。同时用户影响力和用户活跃度并不存在严格的正比关系，影响力高的用户活跃度不一定高。

3) 发布者价值

网络中发布者之间会经常互动，发布者的情感价值观也会相互传递、相互影响。发布者的情感倾向成为了舆情监测的重点。

选择排名前 10 的负向情绪用户（表 4.5），可以看出大多数情感倾向为负值的用户，已经被网站拉黑或者清除，没有被拉黑的用户也可以作为舆情检测的重点对象。

表 4.5 经济领域发布者价值

发布者	情感倾向	发布者	情感倾向
匿名用户 201	-4.66667	匿名用户 409	-2.67500
匿名用户 535	-4.00000	匿名用户 737	-2.40000
匿名用户 643	-3.33333	rollaswold	-1.95000
inkeee	-3.03750	土老巴 2011	-1.80000
匿名用户 250	-2.80000	stupid fox	-1.67867

注：爬取数据时不能获取用户名的用户被标注为匿名用户。

4.2.2 舆情受众

舆情受众主要指所有被舆情影响到的网络用户。衡量舆情受众主要有三个

指标：负面指数、参与度、受众影响力。（指标的计算公式同节 2.3.2）  
本章热点专题的舆情受众指标如表 4.6 所示。

表 4.6 经济领域舆情受众指标

热点专题	负面指数	参与频度	受众影响力
房市与房价	0.50392	0.000456	0.036399
股市与股价	0.59532	0.002117	0.077430
货币与汇率	0.46149	0.005213	0.272111

4.3 专题 1：房市与房价专题

经过 4.1 节的热点专题与热点词分析，选取房市与房价作为经济领域的第一个热点专题。本节将从房市与房价专题的发展简介、专题中网民观点的分类和总结、专题的传播趋势和用户情感以及专题网和个体网分析等方面进行详细叙述。

4.3.1 专题发展简介

房价问题一直都是近几年的社会热点话题。对现今中国房价形势和政策，众说纷纭。一边是喋喋不休宣传“刚性需求”；一边是低收入百姓拼命感叹房价太高；一边在查囤地、增加土地供应；一边不断出现创新高的“地王”。中央和地方政府推出的一些针对性措施，更是使得房价问题成为热点问题的榜首。

房地产大亨潘石屹在 2013 三亚财经国际论坛中这样评价中国房价问题：“整个城市化的过程中间，产业人员的聚集都到城里面来了，可是我们的户籍没有跟上，房地产十几年出现的问题都是成长太快。同时，在互联网上，广大网民对房价问题展开了激烈的讨论”，见表 4.7。

表 4.7 房市与房价专题热帖

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
中国房地产——即将破裂的财富仓库	包海松	2014/6/7	407024	748	146	93	148.3240
银行已经准备好了，地产将大难临头	梦一生 ct	2014/1/7	260127	220	109	77	22.7876
未来房价经济管窥	椰子花生	2014/5/2	380294	612	122	60	109.6510
从房地产到房地惨（从业十年，在聊地产）	我说的你别信	2015/1/11	40125	169	32	36	38.8017

续表

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
从我身边的人的看中国房产今后十年的发展	凡卜	2015/3/18	89711	71	14	18	11.9800

4.3.2 观点分类与观点列表

根据网络数据信息，经过网民在回复过程中的观点挖掘与倾向进行语义的处理，形成了本专题内的核心观点。以论坛中网民的回复为研究对象，通过观点挖掘和语义分析形成本专题内的核心观点，并对其进行分类和详细描述。

1) 观点分类

根据观点涉及的特征词汇、领域以及网民发表的各个观点内容本身之间存在的相关性与差异性，可以将所有评论分为以下五个主题。

主题一：楼盘介绍

此分类观点包含了三个关键特征词：项目、品牌、商业。对应的观点包括：

观点 1. 对于所介绍的楼盘表现出了足够的兴趣。

观点 2. 对于所介绍的楼盘持观望态度。

主题二：房价的涨跌和房屋质量对房价的影响

此分类观点包含了以下三个关键特征词：不动产、房价、土地。对应的观点包括：

观点 3. 房屋的建筑面积质量欠佳必定会影响价格，购房者需自己承担应有的损失。

观点 4. 购房者应该分析清楚广告与现实的差距，避免在购房时被开发商的“花言巧语”所欺骗。

观点 5. 刚性需求增多，房价将会继续涨。

观点 6. 房价是由多种因素影响的，政府的调控才是房价波动的根源。

主题三：房地产投资与投资多元化

此分类观点包含了以下几个关键特征词：市场、价格、房地产、投资。对应的观点包括：

观点 7. 应该理性的多元化投资，房地产的投资仅仅是其中的一部分。

观点 8. 投资房地产简单，收益可观，应该重点投资。

观点 9. 理财投资都是“土豪”的“游戏”，穷人们还是想着先买房子安身吧。

#### 主题四：房屋装修的问题

此分类观点包含了以下几个关键特征词：装修、风格、选择、问题。对应的观点包括：

观点 10. 房屋装修价格之间的讨论。

观点 11. 对于发帖者的装修的效果和时间的评价。

观点 12. 正在向发帖者学习，以便自己在后来装修时少走弯路。

#### 主题五：房屋位置的选择

此分类观点包含了以下几个关键特征词：房子、配套、生活、学校。对应的观点包括：

观点 13. 选择地铁路口、商场周边等交通便利的地方买房。

观点 14. 为了孩子考虑，学区房是买房的不二首选。

观点 15. 价格是买房首要考虑的问题，应该联系自己的实际情况，仔细考虑。

从图 4.2 可以看出，在参与“房地产问题”讨论的网民中，大部分网民关心的是房价涨跌和房屋位置的选择。因为房价和房屋位置是其他三个方面的基础，想要买房的人最关心的也是房价和房子的位置；从房地产投资占有一定的比重可以看出，广大网民的投资意识在迅速提高，同时也有部分网民积累了一定的财富；从楼盘介绍可以看出，开发商的楼盘广告和软文对于网民的购房存在一定的影响；从房屋装修所占比例至少可以看出，在房价波动、房屋位置选择领域活跃的用户有可能并没有购买房屋、房屋装修的环节。

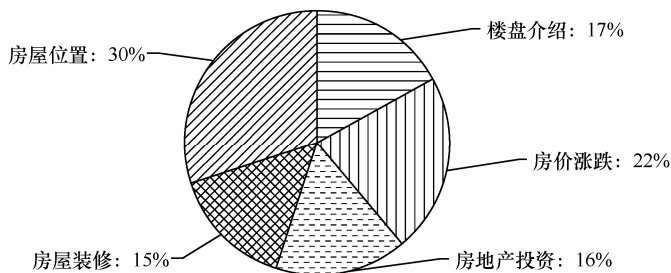


图 4.2 “房市与房价”专题观点分类比例

#### 2) 观点分类详细分析

为了更好地从宏观分析转向微观深入分析，本节对每类观点的网络用户分布以及民意进行了进一步分析，希望反映出他们对这些问题的思考与看法。

### 主题一：楼盘介绍

在进一步的数据（成都恒大御景半岛、成都凯德卓锦万黛以及深圳罗湖金岸商场）分析与研究中，对网民在这个主题内的深入观点进行梳理得知，存在2个细分的子观点如下：

观点1. 12%的评论认为楼盘的广告、软文以及一些网民热捧的楼盘并不值得信赖，自己的实地考察和亲朋好友的切身体会才是影响自己选择楼盘的重要因素。这样的选择结果具有其存在的理由，现在的楼盘广告铺天盖地，夸大其词，确实没有几分的可信度。同时网络中充斥着各类软文，它们与网民的真实想法混淆在一起，导致有些网民只相信自己身边的人，并不打算从网络中获取关于楼盘选择的有效信息。

例如，“如何打造黄金商城商铺？我也想知道”。

观点2. 88%的评论认为楼盘的广告和网民热捧的楼盘还是具有一定的可信度，里面去除广告和软文外，还有大量网民的理性分析，广大网民需要有自己的思考后有所取舍。

例如，“恒大御景半岛的配套绝对一级棒，什么都有”“深圳罗湖金岸商场商铺看起来不错！”“挺喜欢卓锦万黛的园林的，挺高大上的。”

在本主题中，具有相应子观点的评论分布比例如图4.3所示。

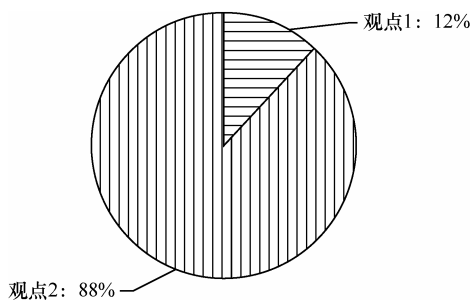


图 4.3 主题一细分子观点比例图

### 主题二：房价的涨跌和房屋质量对房价的影响

在进一步的数据分析与研究中，我们对网民在这个主题内的深入观点进行梳理与分析，其中在本主题下，存在4个细分的子观点，这些细分的子观点所反映的内容如下：

观点1. 40%的评论认为购买到建筑质量有问题的房屋应该自己承担房价下降所造成的损失。大部分网民在BBS中的态度是“事不关己高高挂起”，他们认为购房者在买房前没有理性判断楼盘的虚假广告，由于楼盘降价所造成的损失自己承担。

例如，“就是，不喜欢卖了呗”“主要是买贵了，别的都是假的。”

观点 2. 10%的评论表示买房子时候一定要有理性思维,并对已经购买建筑质量有问题的购房者表示同情。

例如,“主要是买贵了,其实呢,图纸,模型,这些知识广告的宣传手段。不要忘了,最后一句以实物为准”“唉……居不易。”

观点 3. 25%的评论支持房价上涨的帖子分析正确。

例如,“啥也不说了,赶紧买去吧。”

观点 4. 25%的评论表达评论者自己的思想,并不认同房价上涨的观点。

例如,“这种分析水平太次,看一眼不忍直视,走了……”

在本主题中,具有相应子观点的评论分布比例如图 4.4 所示。

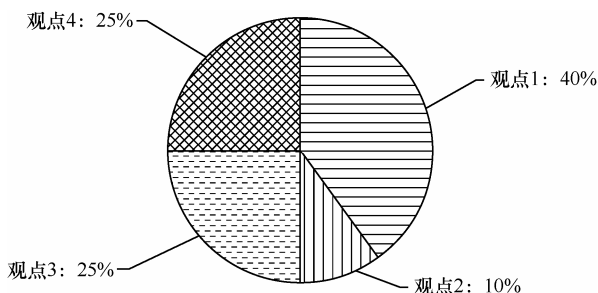


图 4.4 主题二细分子观点比例图

### 主题三：房地产投资与投资多元化

在进一步的数据分析与研究中,对网民在这个主题内的深入观点进行梳理得知存在 3 个细分的子观点,这些细分的子观点所反映的内容如下:

观点 1. 47%的评论认为应该理性的多元化投资,房地产的投资仅仅是其中的一部分。网友对于投资多元化还是比较支持的,股票基金的支持者数量明显高于炒房,而且股票基金等投资相对资金较少,门槛较低;房地产投资门槛高,金额较大,回报周期较长。

例如,“呵呵你真厉害,佩服哦说得不错,有收获,顶一下”“加 QQ 群了,还请楼主多多指导怎么做的都听别人说好做,自己还没尝试呢。”

观点 2. 16%的评论认为投资房地产简单,收益可观,应该重点投资。

例如,“房地产都是有钱人玩的,而且收益不会低……”

观点 3. 37%的评论认为理财投资都是“土豪”的“游戏”,穷人们还是想着先买房子安身吧。部分网友还是比较悲观的,认为房子都没有,还谈什么投资。

这是一种不科学的思想，投资的回报依然可以用来买房，投资意识应该培养。

例如，“所以说贫富差距太大，有钱人和隐形有钱人太多，但普通人和穷人也多。”

在本主题中，具有相应子观点的评论分布比例如图 4.5 所示。

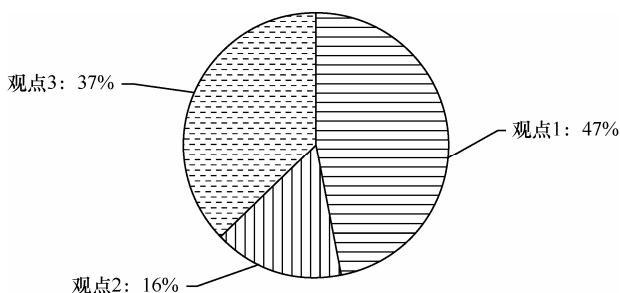


图 4.5 主题三细分子观点比例图

#### 主题四：房屋装修的问题

在进一步的数据分析与研究中，对网民在这个主题内的深入观点进行梳理得知存在 3 个细分的子观点，这些细分的子观点所反映的内容如下：

观点 1. 56% 的评论集中于房屋装修价格之间的讨论，而且发帖者和评论者之间的频繁交互。

例如，“嗯，大概六万吧”“花了大概 10 来万吧，还没仔细的算清楚呐”。

观点 2. 22% 的评论是对于发帖者的装修的效果和时间的评价。

例如，“感觉楼主的房子装修出来会很漂亮”“只是觉得床上不应该吊吊灯，其余的都很不错。”

观点 3. 22% 的评论表示回复者正在向发帖者学习，以便自己在后来装修时少走弯路。

例如，“挺厉害啊！装成这样大概花了多少钱啊”“这个风格我喜欢，你是设计师哇，我房子今年拿，能否谈谈。”

此主题下的评论以正向情感为主，说明谈论装修话题的评论者们都比较积极，应该是拥有自己的房子后，心情整体比较愉悦的原因。大部分装修的话题集中于装修的花费，说明广大网友还是对于金钱最为关注，装修的效果和时间比起花费，热度相对较低。

在本主题中，具有相应子观点的评论分布比例如图 4.6 所示。

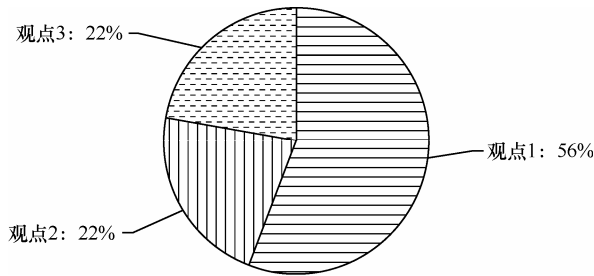


图 4.6 主题四细分子观点比例图

### 主题五：房屋位置的选择

在进一步的数据分析与研究中，对网民在这个主题内的深入观点进行梳理得知存在 3 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 36%的评论认为应该选择地铁路口、商场周边等交通便利的地方买房。

例如，“单身大龄女青年……选七宝地铁房为宜。”

观点 2. 34%的评论表示为了孩子考虑，学区房是买房的不二首选。

例如，“国际学校成趋势了，以后留学越来越方便”“学区房应该不会跌价吧。”

观点 3. 剩下 30%的评论认为价格是买房首要考虑的问题，应该联系自己的实际情况，仔细考虑。

例如，“别墅配国际学校，有钱人玩的东西我还是不要碰。”

对于房屋地理位置的选择，网友的意见较为分散，有支持选购交通便利的楼盘，让自己的生活更轻松、更惬意；有的支持选购学区房的，努力为下一代创造好的条件；还有的支持选购性价比比较高的楼盘，不要给自己带来太大的财务负担。生活环境的不同会导致不同的选择，在房屋地理位置选择的问题上，网友们还是需要结合自身的条件，深思熟虑后再进行选择。

在本主题中，具有相应子观点的评论分布比例如图 4.7 所示。

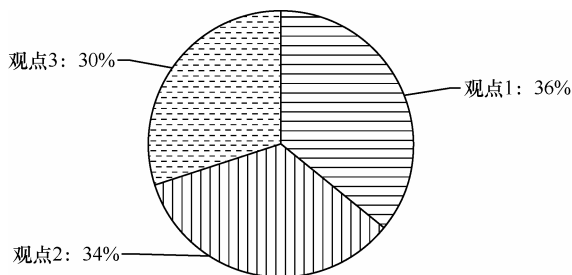


图 4.7 主题五细分子观点比例图



### 3) 观点总结

房地产作为中国经济的支柱性产业，和每个人都有密不可分的关系。从楼盘广告到楼市价格走势，房屋装修、楼盘地理位置分析等话题，可以看出广大网民对于房地产比较关注。但是，楼市价格较高、开发商虚假广告、装修材料劣质等问题，依然充斥在网友的评论中，说明社会中还存在诸多不和谐的因素，导致网友们纷纷在 BBS 中发泄自己的负面感情。希望有关部门可以有效控制房价、加强监管开发商、严惩劣质装修材料生产商，让广大人民群众可以摆脱这些问题的困扰，都可以拥有自己的房子、自己的家、自己幸福美满的生活。

#### 4.3.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

##### 1) 影响力

专题的传播过程中流量是指相关帖子的回复数量，发展峰值出现在舆论讨论最激烈的时间点。很明显，流量与影响力成正比关系，即流量越大影响力也就越大。因此，本节分别对新增回复数和总回复数按照时间进行统计，然后对统计结果进行可视化得到波线图（图 4.8，图 4.9），以便发现专题传播的发展峰值。

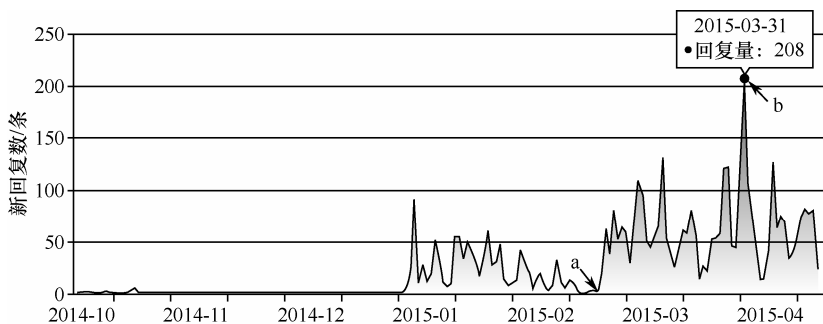


图 4.8 专题日流量时间趋势

点 a: 2015 年 2 月 25 日; 点 b: 2015 年 3 月 31 日

从图 4.8 可以看出，每天对应的新增回复数的变化。在点 a 之前，每天对应的新增回复量有时增加、有时下降，但相对来说，都是比较低，说明这一段时间该事件提出之后并没有引发很大的影响；而在点 a 之后，每日新增量开始瞬间增高，并持续整体上涨趋势，直到在点 b 时达到发展峰值，说明该专题在点 a 之后引发的讨论越来越激烈，在点 b 到达了最激烈的程度。在点 b 之

后引发的讨论稍微下降，但也偶尔也有上涨情况，但是总体来说比点 a 之前讨论量多，说明在点 b 到达一个发展峰值，可能是在该天发生了某个事件，而引发了人们激烈的讨论。

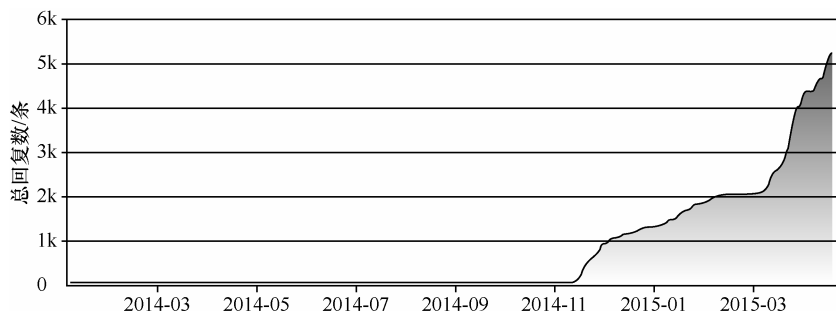


图 4.9 专题总流量时间趋势图

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库，对专题内的回复内容进行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图（图 4.10）和专题情感分类图（图 4.11）。

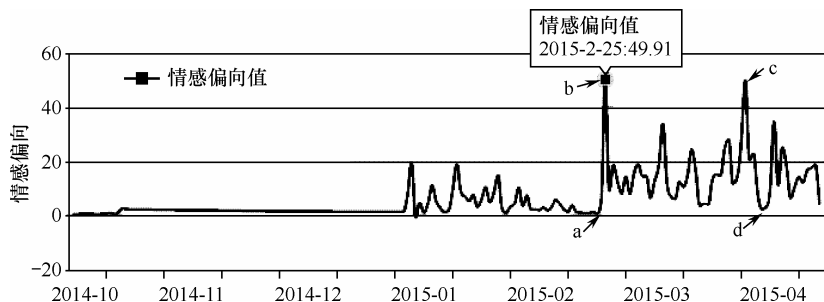


图 4.10 专题情感趋势图

点 a：2015 年 2 月 23 日；点 b：2015 年 2 月 25 日

点 c：2015 年 3 月 31 日；点 d：2015 年 4 月 4 日

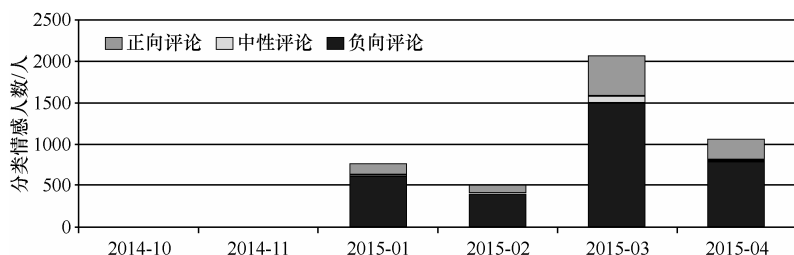


图 4.11 专题情感分类图

从图 4.10 可以看出,有 b 处、c 处两个峰值。在点 b 和点 c 对应日期所对应的情感偏向的情感值比较高,说明这两个日达到了情感的发展峰值;而在点 a 和点 d 等,情感处于较低状态,接着发展到峰值,说明这两点属于情感拐点,必定有一些事情影响了网民的情感。此外还可以看出,在房地产领域网民的情感极易受到影响。结合图 4.8 和图 4.10 可以发现一些规律:①图 4.10 中 2 月 25 日达到情感峰值,图 4.8 中 2 月 24 日之前日增长量比较少,说明该日之前的情感大多数是正值。②图 4.8 中的影响力发展峰值对应的日期情感值也同时达到峰值,说明流量和情感值也有一定的正相关性。从图 4.11 可以看出,在网民的评论中,负面评论占的比重较大,说明现在网络环境下,喷子(一般指在回帖或视频评论的时候肆意谩骂的人)较多,网民发言还是比较随意,大部分网民选择在网络上发泄自己的情绪。

#### 4.3.4 专题用户情感的分析

随着近十年来论坛的兴起,网民数量也急剧上升,网民作为网络舆情的一个非常重要因素,行为和言论会引导着整个舆论的走向。而在网民中那些发言多、回复多的“意见领袖”或者“领军人物”,往往会主导整个舆论的走向。所以说,探索专题下的人物情感可以帮助我们发现该专题下的情感正向用户,同时也可以帮助论坛找出到处散发消极影响的网民,剔除该用户以净化论坛环境。

本节对各专题下的人物情感进行了统计分析;由于活跃度对情感的影响较大,所以本节构建了分别以情感值和活跃度为坐标轴的专题下的用户散点图(图 4.12)。

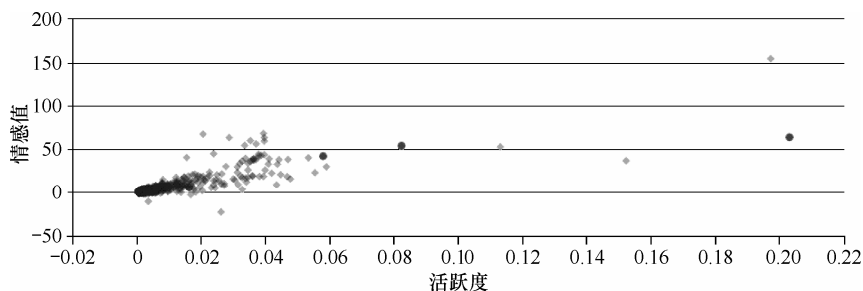


图 4.12 基于活跃度和情感值的用户散点图

圆点代表发帖人;方点代表回帖人

从图 4.12 中可以看出,一般发帖人的情感值都比较高一些。但是,也存一些点不是发帖者而其活跃度较高,说明有两种可能:第一种是该用户作为评论人,评论的内容有又发了一个新的讨论热潮,第二种可能就是属于那些发帖较少,

经常参与那些评论或者原帖回复的那一类人。这些人有一定的回复参与量,但是通常回复的情感极性较低,感情色彩不明显,这些人所属领域性比较明显。

为了进一步进行分析,本节将那些情感值负性较大且活跃度较小的人进行了详细点的描绘,取情感值小于 20、活跃度大于 0.02 的所有用户作图。这类用户代表的是那些在论坛活动积极,但是负面情绪较大的。这些人的活跃度相对较大,他们影响其他用户的可能性也比较大,会给论坛用户灌输一些负面的情绪。因此,我们很容易从这类人群中发现一些核心用户,并且对其进行权限的限制或者屏蔽、删除操作。

从图 4.13 中的点可能是那些传播消极情绪的用户,论坛管理员可以对这些用户进行监察和查询历史,以及更进一步的操作。

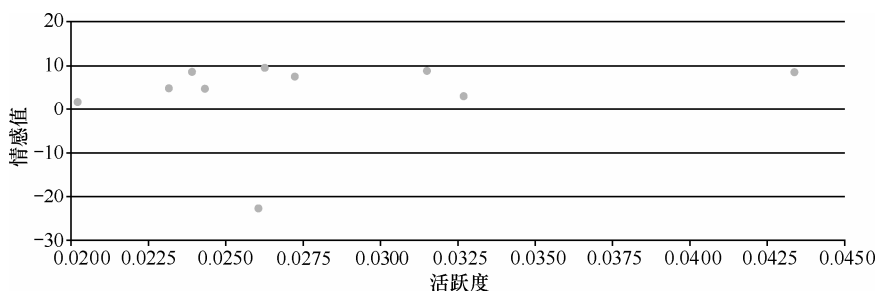


图 4.13 低情感和高活跃度散点图

### 4.3.5 专题网络的构建与分析

专题网络是由该专题下热点事件的发布者、热点事件评论者构建的。该社会网络能够清晰看出热帖的传播过程,发现主题事件传播过程中的重要节点,这些节点就是该主题传播中的意见领袖。此外,通过研究专题网络的结构,能够发现专题的传播模式。专题网络的构建过程包括以下几步:

第一步:提取该主题下的热点事件,通过事件热度指标进行筛选。

第二步:通过热点事件得到原始帖子以及相关评论。

第三步:处理每条评论。查询每条评论预处理中提取出的@论坛用户列表,如果列表为空,建立评论用户 ID 和评论原帖发布者 ID 之间的对应关系;如果列表不为空,建立评论用户 ID 和@的用户 ID 之间的对应关系。

第四步:将所有的对应关系写入 Excel 文件,转存为 csv 格式的文件。

第五步:使用可视化工具 Gephi 进行可视化。

对房市与房价专题构建的专题网络进行可视化,结果见图 4.14。

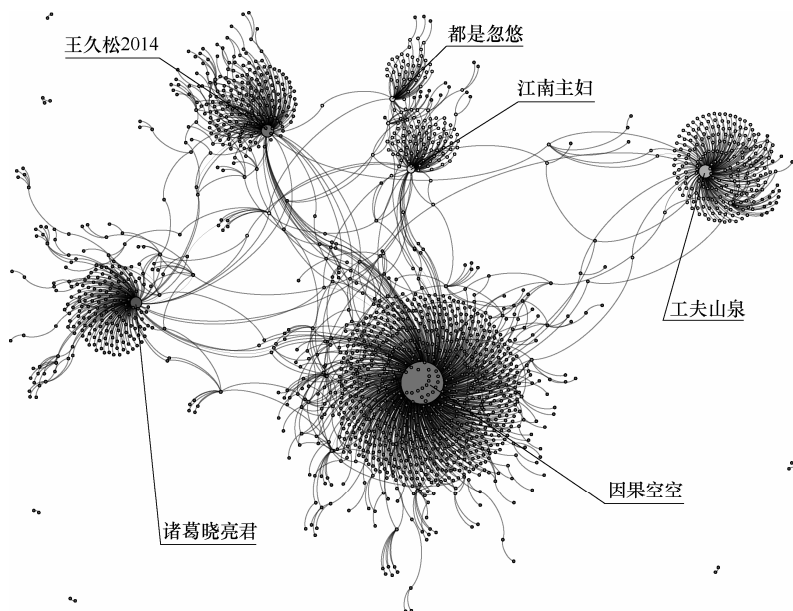


图 4.14 专题传播网络

1) 网络指标概述

本节分别从社会网络的整体指标（表 4.8）、中心性与影响力指标、关联性与“小世界”指标进行了分析，以便发现网络中的特殊节点，以便下一步对个体网演化和传播的分析。

表 4.8 基本网络指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
2.404	0.001	9	0.662	3.662	0.466

平均度：较小。图密度：比较稀疏。模块化指标：划分出 27 个社区，其中 5 个社区规模远远大于其他社区。

2) 中心性与影响力

中心性指的是位于中性或具有中心的性质。中心性主要可以分为三个方面来衡量：点度中心性、中间中心性和接近中心性。每种中心性都有对应的中心度和中心势两个指标。中心度是描述图中某一节点居于某种中心的衡量指标。中心势是描述整个图的紧密程度或一致性，也就是一个图的中心度。

**中心度：**描述图中任意一点在网络中占据的核心性。中心度主要可以分为三个方面来衡量：点度中心度、中间中心度和接近中心度。点度中心度又包括

绝对中心度和相对中心度，绝对中心度就是指该点的度数，相对中心是绝对中心度与图中点的最大可能的度数只比。中间中心度是用来测量社会行动者对资源控制的程度。接近中心度表示一种不受他人控制的程度。

从点度中心度、中间中心度、接近中心度以及它们的相对中心度进行分析中心性和影响力。使用 UCINET 对该网络的中心度指标进行计算，这里取点度中心度排名前 6 的节点，列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和值排名，得到对应的数据表如表 4.9 所示。

表 4.9 中心度指标

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	值排名	绝对值	相对值	值排名	绝对值	相对值	值排名
因果空空	1331	1.240	1	13765	0.416	3	668882	0.272	1732
工夫山泉	458	0.427	2	9390	0.284	5	2897688	0.063	1721
王久松 2014	293	0.273	3	8678	0.262	1	2963188	0.061	1674
诸葛晓亮君	253	0.236	4	5801	0.175	6	337478	0.539	1732
江南主妇	194	0.181	5	5316	0.161	2	2996004	0.061	1629
都是忽悠	73	0.068	6	4898	0.148	8	668882	0.272	1628

从表 4.9 中可以看出，“因果空空”“工夫山泉”等 6 个用户排名前 6 位，说明他们的点度中心度和中间中心度都比较大，排名比较靠前；而接近中心度相对排名比较靠后，说明他们的接近中心度比较小。这说明这几个点在对应的网络中处于中心地位，并且和其他节点的距离都很大。

一个图也具有一定的中心性质。为了与点的中心度相区别，称图的中心性质为“中心势”。中心势指的并不是点的相对重要性而是图的总体整合度或者一致性，其值介于 0~1。很明显完备图的中心势为 0 表示每个点都有相互联系，无所谓中心不中心，星型或辐射型的网络的中心势接近 1。

图 4.14 的点度中心势指标为 0.23%，该指标值较小，表示该图在度数上向某点集中的趋势并不明显；该图的中间中心势指标为 0.42%，该指标值较小，表示该图具有中间节点的程度不是很大，即在整个网络中大部分的节点不需要经过某些中间节点，便可以得到信息。

3) 关联性与“小世界”

网络图中大部分的节点不与彼此邻接，但大部分节点可以从任一其他点经少数几步就可到达。

小世界网络可以反映陌生人由彼此共同认识的人而连接的小世界现象。小世界拥有四个方面的性质：①整个网络巨大，现实世界中包含的人数达到十亿数量级。②网络是稀疏的。③网络是去中心化，即不存在核心点。④整个网络是高度聚类的。

特征途径长度和聚类系数是小世界的两个测量指标。

在网络中，连通任意两个节点的最少边数，就是这两个节点的路径长度，网络中所有节点对的路径长度的平均值，就是该网络的特征途径长度，也称为平均路径长度。网络的凝聚度表示了图中各点之间联络的紧密程度。具体来说，是指在网络中实际存在的线与可能存在的所有线的比值。它用来测量社会网络中行动者之间的联系程度，即整个，值介于0~1。

图 4.14 的特征途径长度为 2.854，凝聚度为 0.005。图 4.14 路径长度的具体分布如表 4.10 所示。

表 4.10 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	2263	0.056	5	3735	0.093
2	18677	0.466	6	785	0.020
3	9301	0.232	7	717	0.018
4	4561	0.114	8	18	0.000

聚类系数是表示一个图形中节点聚集程度的系数。假设某个节点有  $k$  条边，则这  $k$  条边连接的节点 ( $k$  个) 之间最多可能存在的边的条数为  $k(k-1)/2$ ，用实际存在的边数除以最多可能存在的边数得到的分数值，定义为这个节点的聚合系数。所有节点的聚类系数的均值定义为网络的聚合系数。聚类系数是网络的局部特征，反映了相邻两个人之间朋友圈子的重合度，即该节点的朋友之间也是朋友的程度。

图 4.14 的聚类系数为 1.301，节点的聚类系数如表 4.11 所示。

表 4.11 聚类系数指标

用户	聚类系数	点对数	用户	聚类系数	点对数
因果空空	0.000	22003	诸葛晓亮君	0.001	16110
工夫山泉	0.001	22155	江南主妇	0.003	5778
王久松 2014	0.001	18721	都是忽悠	0.015	1378

4.3.6 个体网络分析

在专题领域的整体网分析之后，我们选择了一些传播特性相对较强的用户，又研究基于一个帖子下所有用户的个体传播网络。如图 4.15 所示，选取了“江

南主妇”所发的帖子下对应的所有用户。

由图 4.15 可以计算出个体网络的指标：结构洞、诚实中间人。

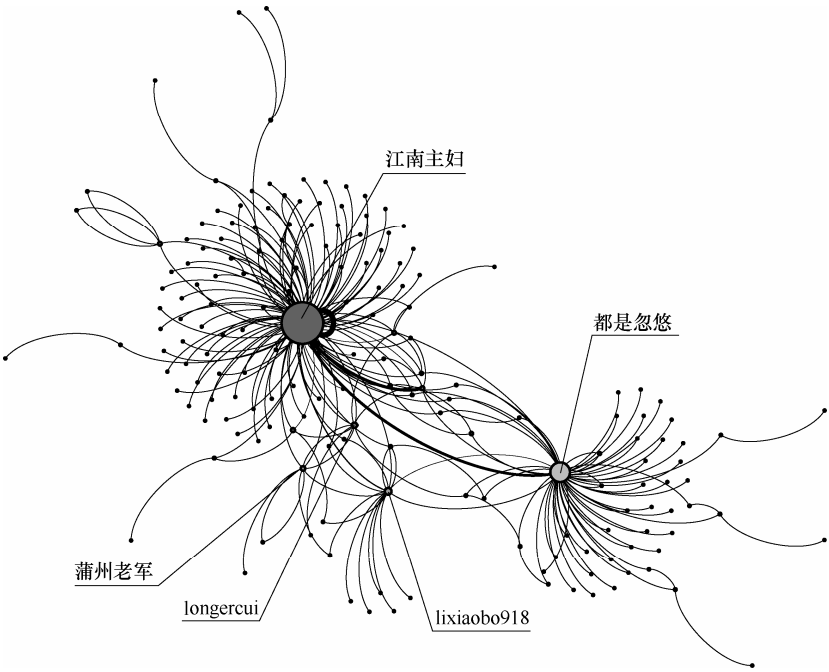


图 4.15 “房市与房价”个体网络

1) 结构洞

一谈到结构洞，至少涉及三个人。对于三个行动者 A、B、C 来说，如果 A 和 B 有关联，B 和 C 有关联，而 A 和 C 无关系的话，我们就说这种结构是一种结构洞，或者说在 A 和 C 之间存在一个结构洞。结构洞的存在使得 B 处于中间人地位，B 因而可以控制资源的传递等。在管理者的社会资本中，弱关系和结构洞占据重要地位。

这里列举节点度数排名前 5 位的用户的结构洞指标，形成相应的数据指标表如表 4.12 所示。

表 4.12 结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
江南主妇	109	87.855	0.806	0.335	0.889
都是忽悠	53	51.330	0.968	0.130	0.683
lixiaobo918	13	11.716	0.901	0.439	0.584
longercui	8	5.727	0.716	0.611	0.358
蒲州老军	6	4.989	0.831	0.538	0.322



从表 4.12 中可以看出,“江南主妇”所对应的有效规模数量值最大(87.855),说明“江南主妇”居于网络的核心,其次是“都是忽悠”。效率值最高(0.968)的为“都是忽悠”,说明该用户在社会网络中受到的限制最大。“longercui”对应的限制度值最大(0.611),说明该用户在网络中受到的限制最大。

## 2) 诚实中间人

从表 4.13 中可以看出,“江南主妇”代表的节点作为中介的次数为 17 次,共联络了 136 对关系,其中 126 对关系属于 HBIO 类,8 对关系属于 HBIO1 类,2 对关系属于 HBIO2 类。这说明大部分节点属于纯中介,说明参与回复主贴的人物之间的关系比较弱;同时也代表大部分回复是针对主贴回复的,而只有一小部分回复针对的是评论用户评论的。

表 4.13 诚实中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
江南主妇	17	136	126	8	2
都是忽悠	4	6	5	0	1
蒲州老军	3	3	2	0	1
人愚的等傻	3	3	3	0	0

HBIO 代表纯中介, HBIO1 代表弱中介, HBIO2 代表非中介。

从图 4.15 中可以看出,“江南主妇”发表的帖子后,大量用户参与的讨论,在回复的用户中,“都是忽悠”的用户发表的评论引起了网民大量回复。说明该用户在网络传播中可能是一个意见领袖,他的主观意见会影响或者左右其他用户的情感倾向。

## 4.4 专题 2: 股市与股票

即便过去一年经历了大幅上涨,中国股市相对于国内生产总值(GDP)的比重仍无法企及美国或日本。对比股市与银行信贷及债务规模,中国经济仍然极大依赖银行体系。中国股市何去何从对改革进程至关重要。通过股市、债市直接融资,对降低信贷成本、克服银行体系的政治经济问题,将资金引向高潜力企业至关重要。本节对股市问题进行舆论的挖掘,找出股市方面的舆论,以及股民所关注的问题。

### 4.4.1 专题发展简介

随着时代的发展,人们的生活水平越来越高,人们的重心从原来的温饱问

题,接着转向购买和投放股票。据不完全统计,在城市人口中大约 60%都有买股票的经验,同时在农村人口中买股人也占了一个很大的比例。所以,股票问题成了一个普遍性的问题;同时,由于股票的可变因素比较多,主要受国家政策的影响,国家政策往往会主导股票走势;此外,社会舆论也对股票影响非常大,一些公司的负面报道往往会导致股票大跌。

近二十年来,大部分公司上市后都公开发行自己的股票,并且公开发行股票发放量往往占总股票的 50%以上。由此可见,一个公司股票的走势往往会决定一个公司的效益。对于个体购买股票者来说,一个好的股票导向往往会直接影响个人的盈利。所以,探索股票方面的舆论问题具有深远的影响。

表 4.14 是股市与股票专题的几大主要热帖。在帖子的专题传播过程中,大量的网民针对所发表帖子的观点进行了回复,其中也产生了一些新的且具有一定影响力的观点,并引发了网民的关注和讨论,尽管新观点基本还是发帖者的观点延伸,体现了网友讨论的内聚性和聚焦性。

表 4.14 股市与股票专题热帖

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
自然法则之时空纵横 A 股、恒生、美股、美指、黄金、外汇趋势预测大家谈	chenkx214	2014/11/11	282946	1654	471	356	113.3240
职业股民,靠股市生存的人,有什么要问的,尽管问	稳健 LJ	2014/06/04	57601	398	83	74	84.4565
淡定炒股,悠然生活	suifengyuanshi	2015/01/15	16894	236	55	32	67.6815
2014-5 美指绝境反击王者归来,金银回光返照先扬后抑	chenkx214	2014/01/06	1956116	22	41	3	5.9770
3 月:中小行情中线顶尾声,蓝筹跌至 5 月见中线底	罗布次	2015/03/04	5517	36	14	5	10.0522
自然法则之时空纵横 A 股、恒生、美股、美指、黄金、外汇趋势预测大家谈	chenkx214	2014/11/11	282946	1654	471	356	113.3240

4.4.2 达式观点分类与观点列表

通过上述的网络数据信息,分析此专题中网民在帖子中发表的全部言论,针对网民在回复过程中的观点挖掘与倾向进行语义的处理,形成了本专题内的核心观点。

### 1) 观点分类

将该专题下的所有帖子，经过文本清洗、语义分析、观点聚类这三个核心步骤，提取到3个核心观点。这3个观点占整个用户分布的比例数据如图4.16所示。

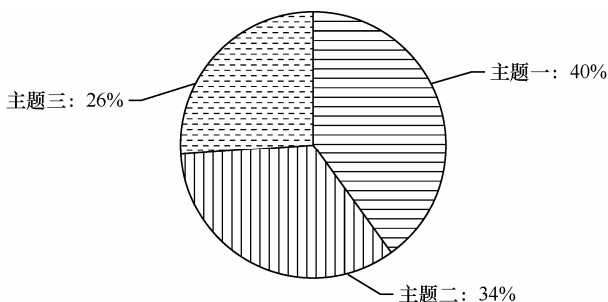


图 4.16 “股市与股票”专题观点分类比例图

**主题一：**创业板估计会继续维持相对于主板的强势格局。持有该观点的人数占总人数的40%。

**主题二：**持股的不要慌，空仓的不要忙。持有该观点的人数占总人数的34%。

**主题三：**操作难度将是越来越大，闭上眼买票赚钱已不再来，踩中跌停就更惨。持有该观点的人数占总人数的26%。

从上述三个主题的热门观点可以总结出，越来越多的股民关注的并不是专业的股市新闻，而是股市的方法，炒股的心态，以及他人分享的经验。由此可以发现，现在中国大部分股民不是那么专业，股市的不懂行情的用户比较多，说明大家期望从别人经验中来选择如何炒股或者买哪只股票。

通过网民已发表的所有评论内容，统计分析相关观点和讨论所涉及的热点词汇、重要领域，通过针对文本的分词、去噪、清洗、过滤等操作，聚合出关于“股市问题”专题的发展演化过程中存在的10个观点特征词汇：创业板、强势、格局、持股、空仓、操作、股票、赚钱、跌停、股票。

### 2) 观点分类详细分析

为了更好地对相应的观点分类进行深入的分析，本节对每个分类主题内的网络用户分布以及网络民意进行了分类处理，希望通过网络用户真实反馈的信息来反映他们对这一些问题的思考与建议。

### 主题一：创业板估计会继续维持相对于主板的强势格局

在进一步的数据分析与研究中，通过对本主题内容的深入观点进行梳理与分析，发现存在 3 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 51%的人认为创业板将会继续维持强势。任何一个新事物，在其成熟之前都有一个曲折的过程。创业板也是，有些网友认为创业板将来会继续保持强势。

例如，网友“盛东辉”说：“不知不觉进入两会期间，以金融为代表的沪指蓝筹进入第二波调整。和预计有些出入的是这波反弹没有创出 3406 后的新高，估计这也和三大金融板块之间的涨跌节奏有一些时间差有关。而中小板创业板中得环保、网络、医药、纺织等板块因两会预期题材或补涨等各种原因开始一些表现。从时间上看，金融的调整周期至少还要持续两周之间，而这段期间创业板估计会继续维持相对于主板的强势格局，之后在两会结束后完成创业板和蓝筹的再一次行情转换。现在就是度过沪指第二波调整的艰难时期，预计的时间上面已经提过，至于空间方面，沪指可能会再次跌破 60 日线回踩 3100 附近，然后在牛市不再的舆论恐慌中彻底完成突破 2009 年顶部 3478 的准备。”

观点 2. 27%的人认为创业板将来会被终结。从总体上看，前一阶段二板市大多经历了创建初期的辉煌，但基本在 20 世纪 90 年代中期以失败而告终。所以，以史为鉴，将来创业板也会被终结。

例如，网友“沙地鸵鸟”说：“今日创业板的泡沫远比 2007 年上证 6000 点大关。既是泡沫，破裂在所难免。何时破？怎么破？需要的就是压垮骆驼的最后一根稻草。2007 年，中石油就是压垮大盘这最后一根稻草。2015 年，谁来终结几近疯狂的创业板？毫无疑问，非超级大盘蓝筹科技莫属。”

观点 3. 22%的人认为创业板很容易受到外界影响。这部分认为创业板既不会维持强势也不会被终结，它属于不稳定，也许会受影响，但是不会被终结。

例如，网友“股票投资 1898”说：“周五两市开盘涨跌不一，沪指小幅低开随后冲高，全天维持窄幅震荡，临近尾盘出现震荡走低，沪指报 3241 点，深成指小幅高开，随后冲高回落，全天弱势高开低走！而创业板受唱空影响大跌，失守 2000 点。全天跌幅 3.15%。创业板报 1951 点坚守十日线上方！周五大盘和往日没有什么不同，所不同的是现在唱空创业板的人越来越多，创业板周五放量杀穿 5 日均线，不过目前依然在 10 日均线以上，笔者观点是，创业板短线受多种负面消息影响短线有望调整，但是最后的头部尚没有到来！调整之后将进一步上攻！盘面上来看：高铁概念中国南北受国务院国有资产监督管理委员会原则同意此次合并利好消息刺激，周五再度延续强势！而有色、煤炭、铁路基建、稀土、迪士尼、银行、等板块涨幅居前；题材板块周五成

为杀跌对象，智能手机、谷歌眼镜、智能家居、物联网、网络安全、医疗器械、在线教育、互联医疗、电子支付、计算机等众多题材板块纷纷回落跌幅居前。周五市场权重护盘，创业板指大幅跳水，让市场跷跷板效应再现。这是否意味着风格将再次切换，创业板指将牛途末路呢？我们认为，新一批 IPO 冻结资金甚至有可能高达 3 万亿元左右，同时 GDP 的主动下调也制约着蓝筹股的业绩，资金面与业绩因素无疑让权重股难有持续行情；而小盘题材方面，管理层降低 GDP 增速，意在中国经济转型与升级，新兴产业有望成为中国经济新引擎，大格局上小盘股已符合方向，另外小盘股年报业绩亮丽、容积较小，因此注册制没推出来之前，创业板炒作不会一下子就结束。而是调整之后将进一步上攻！虽然创业板周五出现调整，但其并不会马上见顶，中小市值个股仍将是后市行情的主角。尤其是前阶段表现突出的，随时有梅开二度的可能！因此，积极把握创业板中的互联网，医药，业绩高增长个股机会！具体操作见每日盘中直播！”

综上，相关详细子观点的人群分布比例如图 4.17 所示。

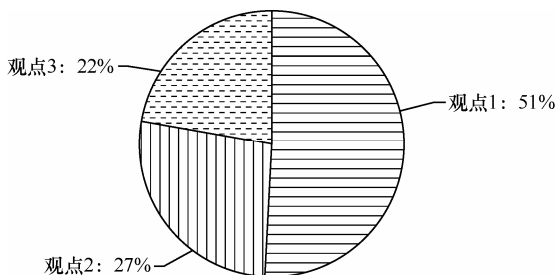


图 4.17 主题一细分子观点比例图

## 主题二：持股的不要慌，空仓的不要忙

通过对本主题内容的深入观点进行梳理与分析，发现存在 2 个主要的细分子观点。这些细分子观点所反映的内容如下。

观点 1. 37%的网友认为炒股宁可错过，绝不做错。大多股民对炒股对股市的认知不同，有些人认为一次投股投错影响不大，往往正是这种想法导致以后股市的认知。

例如，网友信“股动人心 13”说：“2006 年抓到 046 泛海控股，获利 200%（借房地产调控，我黯然吃进，一鸣惊人）；2007 年抓到 0952 广济药业，获利 700%（牛股满天飞，故其他获利牛股就不一一指出）财赋操盘理念：“决心+信心+狠心”，财赋操盘风格：短线为主，快进快出；中线为辅，

精选个股；财赋操盘纪律：止盈止损，保住本金；宁可错过，绝不做错”

观点 2. 63%的网友认为持股不要慌，当今股民持股时。长时间手中持有股票，则会感觉到坐立不安，往往这个时候，股民更需要淡定，不要急于出手。

例如，网友“whw3157”说：“大盘今天的走势还是在对前期高点处堆积的套牢压力进行消化。一般两会期间，由于政策的不确定性，主力一般都是采取来回上下拉锯的策略控制大盘的走势与节奏，几乎每年两会期间的初春行情都是如此。因此在操作上要少安毋躁，保持耐心。持股的不要慌，空仓的不要忙。后市大盘再经过蓄势整理后会开始进入新的上升通道。现在大盘的技术面已经调整的比较充分，这一点无论从均线系统，K 线系统，以及量价关系能上都可以看得很清楚。所以，保持一份耐心再加上一份信心，相信今年的股市一定会带给你一份开心的。”

综上，在本分类观点中，持有相关子观点的人群分布比例如图 4.18 所示。

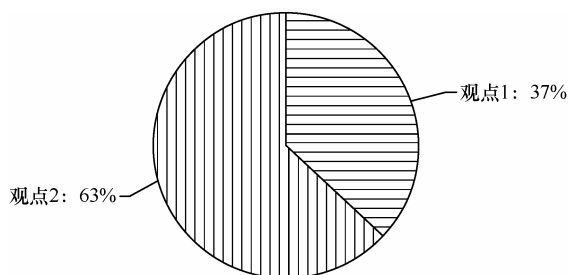


图 4.18 主题二细分子观点比例图

**主题三：操作难度将越来越大,闭上眼买票赚钱已不再来,踩中跌停就更惨**

通过对本主题内容的深入观点进行梳理与分析，大部分网友认为现在炒股是越来越难了，随着股市的暴涨，大家都会有秩序的选择咨询一些股票信息机构，而不是以前那种盲目的购买股票。所以，现在买股票越来越难了，如果不小心猜中跌停，就会赔的更多。

例如，网友“提着电脑满街跑”说：“今日大盘呈现小幅下跌之格局。早盘以 3248 点平开开盘，小幅上涨至 3266 点回调，震荡下行至午市 3234 点，报收于 3241 点，全天下跌 7 个点。成交量较上交易日减少一成，有 3283 个亿。下跌家数远远多于上涨个股，沪深有 717 只上涨，55 只涨停，103 只涨幅 5% 以上。有 1618 只下跌，4 只跌停，170 只下跌超过 5%。从今日盘面观察分析：窄幅震荡、尾市跳水，个股普跌、跌停潮已显，风险已来临，预计

下周还将惯性下跌。今日日K线为小阴十字星，高点触及10日均线回调，低点接近30均线，与昨日小阴线并列阴线，并不是止跌信号。从近期分析：大盘在3400点一线，形成三重顶。在3200~3400点堆积了大量的套牢盘，这一波反弹至3336点回调，已形成小幅下行通道，并无止跌迹象，后市将继续下跌探底，支撑在3049点一线。从日均线系统分析：5日、10日均线已被跌破，5日、20日、30日均线形成空头排列，60日、120日、250日均线依然形成多头排列。下周5日均线将下穿10日均线，在3248点形成压力，支撑在60日均线一线。从波段线画出：压力在3406点，支撑在3212点一线。综合分析：下周还将下跌考验3221点一线的支撑力度，一旦跌破将加速下跌。操作上：从目前情况分析，每月只能做一次，出新股发行消息时做空，耐心等待新股结束时做多，以此循环。现在最好空仓休息到下周五尾市，再选股操作。大盘与个股形成冰火两重天，别管大盘，看住手中个股见顶或转弱就出掉。我是死多头，是以技术面分析选股、短线操作。过山车行情将是常态化，几乎天天都会出现。个股将是两极分化，齐涨齐跌现象已消失。操作难度将是越来越大，闭上眼买票赚钱已不再来，踩中跌停就更惨。”

### 3) 观点总结

从以上观点可以看出，股市领域方面人们关注的不是具体的股票的增长或者跌落，而是大部分关注点移向了炒股的技巧和心态上面，好的技巧和心态会帮助股民更好地适应股票市场，保证在股票市场中有节度的买股和抛股，使股票市场稳定。

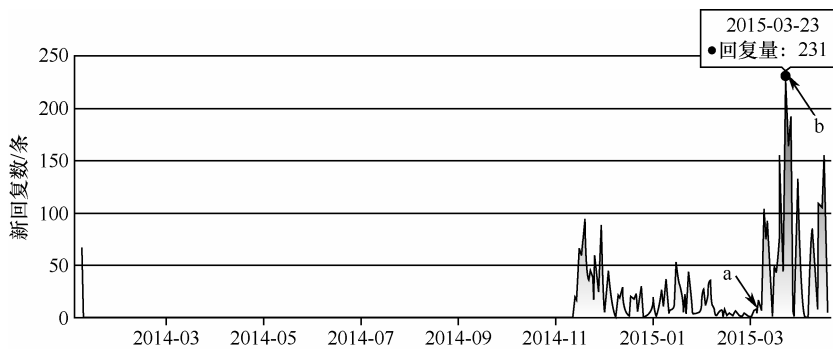
人们的观点重点集中在创业板股票上。在中国发展创业板市场是为了给中小企业提供更方便的融资渠道，为风险资本营造一个正常的退出机制。同时，这也是中国调整产业结构、推进经济改革的重要手段。二板市场和主板市场的投资对象和风险承受能力是不相同的，通常情况下二者不会相互影响。而且由于它们内在的联系，反而会促进主板市场的进一步发展壮大。对投资者来说，创业板市场的风险要比主板市场高得多。当然，回报可能也会大得多。所以，大家对创业板的关注度会比较高，期待创业板持续强势。

#### 4.4.3 专题传播趋势

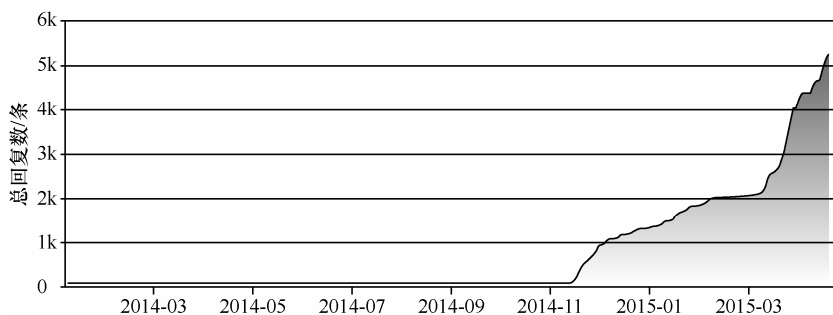
本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

### 1) 影响力

专题的传播过程中流量是指相关帖子的回复数量，发展峰值出现在舆论讨论最激烈的时间点。很明显，流量与影响力成正比关系，即流量越大影响力也就越大。因此，本节分别对新增回复数和总回复数按照时间进行统计，然后对统计结果进行可视化得到波线图（图 4.19，图 4.20），以便发现专题传播的发展峰值。



点 a: 2015 年 3 月 5 日; 点 b: 2015 年 3 月 23 日



从图 4.19 可以看出，在点 a 之前，每天对应的新增回复量有时增加，有时下降，但相对来说，都是比较低，说明这一段时间该事件提出来之后并没有引发很大的影响；而在 a 之后，每日新增量开始瞬间增高，并持持续上涨趋势，在点 b 达到了发展峰值，说明该专题在当点 a 后引发的讨论越来越激烈，在点 b 到达了最激烈。在点 b 之后引发的讨论稍微下降，但是偶尔也有上涨趋势，



但是总体来说比点 a 之前讨论量增加,说明在点 b 到达一个发展峰值,可能是在该天发生了某个事件,而引发了人们激烈的讨论。

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库,对专题内的回复内容进行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图(图 4.21)和专题情感分类图(图 4.22)。

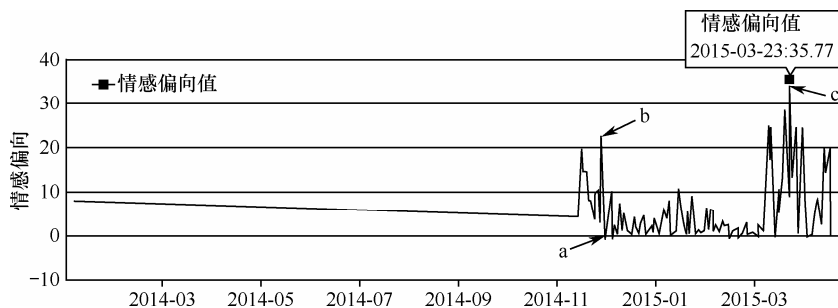


图 4.21 专题情感趋势图

点 a: 2014 年 11 月 27 日; 点 b: 2014 年 11 月 28 日; 点 c: 2015 年 3 月 23 日

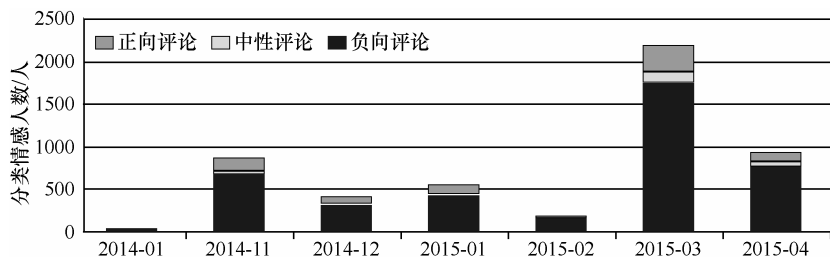


图 4.22 专题情感分类图

从图 4.21 可以看出发展峰值和发展拐点。例如,点 b 和点 c 对应的情感偏向的情感值比较高,说明这两个日达到了情感的发展峰值,而在点 a,情感达到较低的点,接着达到峰值,说明这写点属于情感拐点,必定有一些事情影响了网民的情感。除此之外,还可以看出在股票领域网民的情感极易受到影响。从图 4.22 可以看出在网民的评论中,负面评论占的比重较大,说明现在网络环境下,喷子较多,网民发言还是比较随意,大部分网民选择在网上发泄自己的情绪。

#### 4.4.4 专题用户情感的分析

随着近十年来论坛的兴起,网民数量也急剧上升,网民作为网络舆情的一个非常重要因素,行为和言论会引导着整个舆论的走向。而在网民中那些发言多、回复多的“意见领袖”或者“领军人物”,往往会主导整个舆论的走向。所以说,探索专题下的人物的情感可以帮助我们发现该专题下的情感正向用户,同时也可以帮助论坛找出到处散发消极影响的网民,剔除该用户以净化论坛环境。

本节对各专题下的人物情感进行了统计分析;由于活跃度对情感的影响较大,所以本节构建了分别以情感值和活跃度为坐标轴的专题下的用户散点图(图 4.23)。

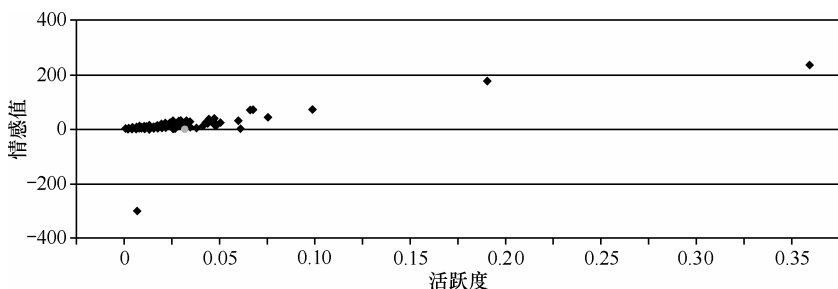


图 4.23 基于活跃度和情感值的用户散点图

圆点代表发帖人,方点代表回帖人

从图 4.23 中可以看出,有一个发帖者的情感值非常大,超过了 300,这是由于该作者发表的帖子是正面的,且引起了广泛的正面讨论;但该用户的活跃度并不是很大,说明该用户在论坛中发帖和回复的频率并不高,可能只是偶然间发的一篇帖子引起了网友的广泛关注。其他的发帖者的情感值和活跃度都相对不高。

为了进一步进行分析,本节将那些情感值负性较大且活跃度较小的人进行了详细点的描绘,取情感值小于 20,活跃度大于 0.02 的所有用户,作图(图 4.24)。这类用户代表的是那些在论坛活动积极,但是负面情绪较大的。这些人的活跃度相对较大,他们影响其他用户的可能性也比较大,会给论坛用户灌输一些负面的情绪。因此,我们很容易从这类人群中发现一些核心用户,并且对其进行权限的限制或者屏蔽、删除操作。

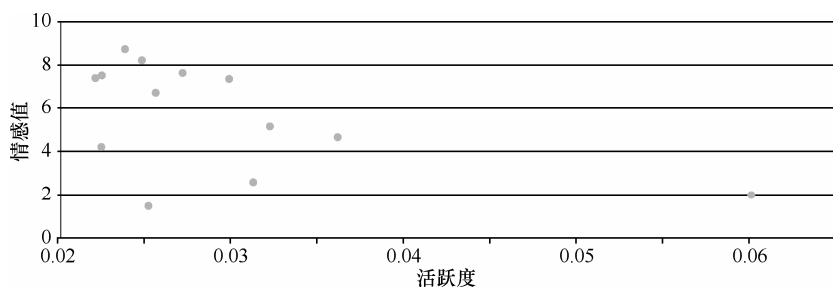


图 4.24 低情感和高活跃度散点图

图 4.24 有两个点（箭头所指的）比较特殊，第一个点，是用户“vspqq”，该用户活跃度比较大，但他的情感属于极低的，这类用户极有可能是负面情绪的发起者，论坛管理员应该采取相应的操作；第二个点，该用户活跃度非常高，比其他用户高出很多，但其情感值却紧紧只有 5（极低的），这类用户极易被利用，发展成为消极情绪的发起者，应予以关注。

#### 4.4.5 专题网络的构建与分析

构建专题网络的方法见节 4.3.5。构建出股市与股票的专题网络，进行可视化，结果见图 4.25。

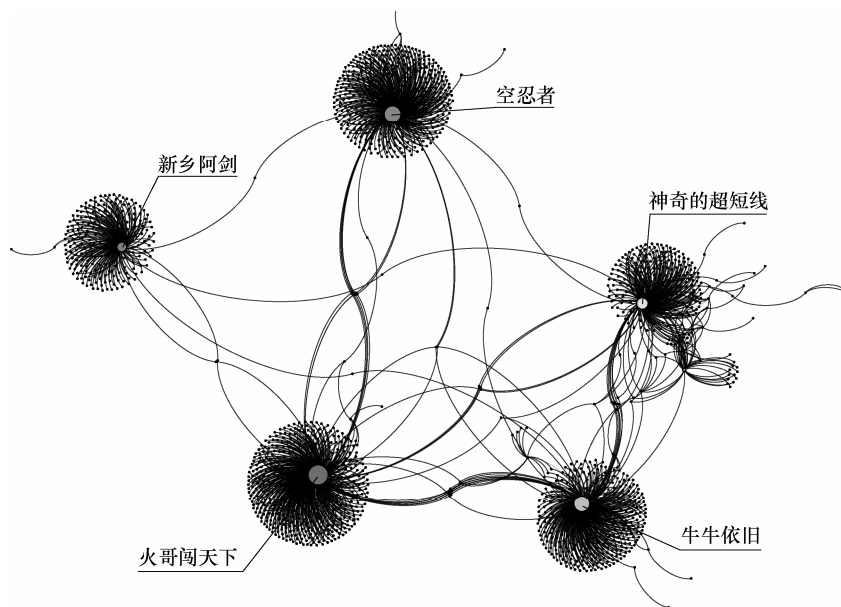


图 4.25 专题传播网络

### 1) 网络指标概述

本节分别从社会网络的整体性指标（表 4.15）、中心性与影响力指标、关联系与“小世界”指标进行了分析，以便发现网络中的特殊节点，以便分析个体网络演化与传播。

表 4.15 整体性网络指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
2.182	0.002	8	0.786	3.571	0.585

平均度：较小；图密度：较小，该传播网络图比较稀疏；网络直径：较长，该网络的传播较远；模块化指标：划分出 8 个社区，且划分的质量较高；平均路径长度：长度较大，网络传播较广；平均聚类系数：较小。

### 2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算，这里取点度中心度排名前 5 的节点，列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名，得到对应的数据如表 4.16 所示。

表 4.16 中心度指标

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	值排名	绝对值	相对值	值排名	绝对值	相对值	值排名
空忍者	1301	0.207	1	185332	7.137	3	536521	0.300	1443
新乡阿剑	1123	0.179	2	67581	2.602	5	241255	0.660	1598
神奇的超短线	1054	0.168	3	146872	5.656	4	536367	0.301	1444
牛牛依旧	997	0.159	4	188096	7.243	2	535458	0.301	1505
火哥闯天下	957	0.152	5	317102	12.211	1	535312	0.301	1399

从表 4.16 中可以看出，“火哥闯天下”的中间中心度远远大于其他节点，排在第一位，说明许多节点都需要经过该节点才能建立联系并获取信息。“牛牛依旧”“空忍者”“神奇的超短线”“新乡阿剑”的中间中心度也相对比较大，排名分别列在 2~5 位，这 4 个节点也各自控制了一部分资源，与图 4.25 中的网络图整体趋势一致。“空忍者”的点度中心度排名第一，与该节点直接相连的节点最多。表 4.16 中列出的 5 个节点接近中心性的值都比较小，说明这几个节点受其他节点的限制或影响很小。

图 4.25 的点度中心势指标为 0.21%，该指标值很小，表示该图在度数上向

某点集中的趋势很不明显；该图的中间中心势指标为 12.19%，该指标值较大，表示该图具有中间节点的程度比较大，即在整个网络中有相当部分的节点需要经过某些中间节点，才能得到信息。

3) 关联性与“小世界”

图 4.25 的特征途径长度为 2.854，凝聚度为 0.005。该图路径长度的具体分布如表 4.17 所示。

表 4.17 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	1985	0.004	5	14263	0.031
2	110601	0.237	6	91045	0.195
3	10716	0.023	7	7311	0.016
4	229643	0.492	8	1077	0.002

图 4.14 的聚类系数为 1.301，节点的聚类系数如表 4.18 所示。

表 4.18 聚类系数指标

用户	聚类系数	点对数
空忍者	0.000	69378
新乡阿剑	0.000	20100
神奇的超短线	0.003	29646
牛牛依旧	0.000	60378
火哥闯天下	0.000	101025

4.4.6 个体网络分析

在按照之前的整体网分析之后，我们从中选取那些节点度数相对较大、传播性比较好的帖子，做了某帖子下参与的所有用户的个体网络。本节选取的是用户“神奇的超短线”发表的一篇帖子，做可视图（图 4.26），对它进行指标分析，并描述传播路线。

1) 结构洞

这里列举节点度数排名前 5 位的用户的结构洞指标，形成相应的数据指标表（表 4.19）。

表 4.19 结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
神奇的超短线	245	177.836	0.726	0.481	0.871
先人 jzy	41	22.926	0.559	0.819	0.829
jiangxinbao520	9	4.591	0.510	1.173	0.605
KSAJHDA	8	5.727	0.716	0.611	0.358
先人 jzy2015	7	2.625	0.375	1.006	0.789

从表 4.19 中可以看出,“神奇的超短线”的节点度数、有效规模和等级度值均为最大,其余用户则远远小于该用户,这说明“神奇的超短线”居于网络最为核心的位置,而且该节点的效率最高,限制度最小,表明该节点在网络中的行动最高效,受到网络中其他节点的限制比较小。“jiangxinbao520”对应的效率值较小,限制度较大,说明该节点在网络中受到其他节点的限制比较大,行动力较弱。“先人 jzy”的等级度比“神奇的超短线”的值小,又远大于其他节点,位于网络次中心的位置。

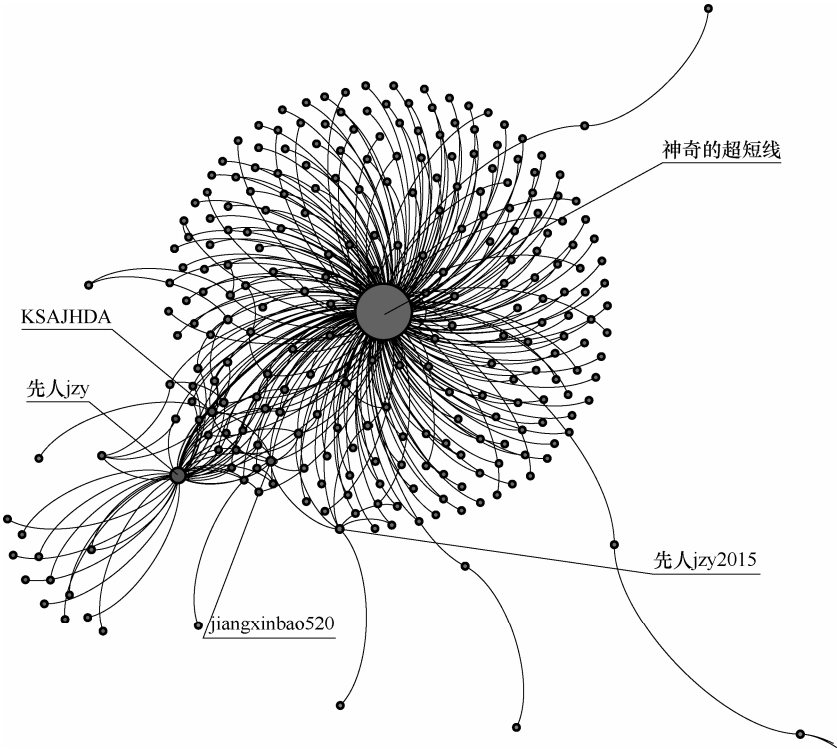


图 4.26 “股市与股票”个体网络

## 2) 诚实中间人

这里列出部分代表性用户节点的诚实中间人指标，形成相应的数据指标表如表 4.20 所示。

表 4.20 诚实中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
神奇的超短线	32	496	229	5	6
KSAJHDA	4	6	2	1	3
先人 jzy	9	36	27	3	6
jiangxinbao520	3	3	0	0	3

HBIO 代表纯中介，HBIO1 代表弱中介，HBIO2 代表非中介。

从表 4.20 中可以看出，“神奇的超短线”代表的节点共联络了 496 对关系，该节点在其中 229 对关系中作为纯中介，在其中 6 对关系中作为弱中介，在其中 6 对关系中作为非中介。“先人 jzy”“KSAJHDA”“jiangxinbao520”等节点也连接了少数节点对，在其中充当了中介。可以看出，“荣耀与卑微”发表了帖子后，大量用户参与的讨论。在回复的用户中，“先人 jzy”发表的评论引起了部分网民的讨论，该节点在网络传播中可能会成为一个意见领袖，可能会影响其他用户的思考和回复。

## 4.5 专题 3：货币与汇率

货币本质上是一种所有者与市场关于交换权的契约，根本上是所有者相互之间的约定。吾以吾之所有予市场，换吾之所需，货币就是这一过程的约定，它反映的是个体与社会的经济协作关系。它实质上就是等价交换的一个过程，当各国之间进行贸易时，由于货币的不同，会有一个价值的衡量，这就是我们所说的汇率。美元一直是世界上最流通的货币，美元的贬值与升值影响国际贸易的走势。随着中国的强大，人民币也逐渐国家化。美元并不是唯一的国际性货币。因此，探索汇率方面问题，具有十分重要的意义。

### 4.5.1 专题发展简介

“汇率”亦称“汇价”，即一国货币兑换另一国货币的比率，是一种货币表示的另一种货币的价格。人民币汇率就是人民币兑换另一国货币的比率。

2005 年 7 月 21 日，人民银行宣布，经国务院批准，人民币汇率改为参考一篮子货币，汇率改为 1 美元兑 8.11 元人民币，变相升值 2%，并且不再与美

元挂钩。自此时间点开始，汇率领域频发争议事件。

随着美国参院通过涉人民币汇率法案的影响，中美两国有关汇率的口水仗激战犹酣。这使得汇率问题不仅成广播电视报纸网络的热点之一，也成了街头巷尾寻常百姓的话题。在一个个专家学者粉墨登场慷慨激昂之后，一个个网民早已被撩拨得热血沸腾义愤填膺。

2014年12月6日，“stupidwz”在天涯论坛上发表了一篇关于名为“俄罗斯，普京，卢布”的帖子，引起了广大网名的关注（表4.21）。随后在2015年2月24日，“09blizzed01”发表的“2015大时代的序幕已缓缓拉开”成为了新的热帖（表4.21）。在帖子的专题传播过程中，大量的网民针对帖子所表达的观点进行了回复，其中也产生了一些颇具影响力的新观点，并引发了网名的关注的讨论。

表 4.21 货币与汇率专题热帖

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
俄罗斯，普京，卢布	stupidwz	2014/12/06	1189006	6168	2093	1419	191.985
经济的癌变-2014中国实体经济走向末日	生于 0715	2014/01/07	1065222	1593	579	216	18.8645
2015 大时代的序幕已缓缓拉开	09blizzed01	2015/02/24	402162	1127	288	175	47.782
谈谈美指、黄金、人民币	椰子花生	2014/10/16	90370	539	50	60	109.651
不作死则不会死，谈谈亚投行	空空儿 2016ABC	2015/03/18	127918	395	98	90	24.4033

4.5.2 观点分类与观点列表

通过上述的网络数据信息，针对网民在回复过程中的观点挖掘与倾向进行语义处理，形成了本专题内的核心观点。

1) 观点分类

根据观点涉及的特征词汇和观点所涉及的领域，以及网民发表的各个观点内容本身之间存在的相关性与差异性，将所有评论分为三个主题。每个主题对应的观点人数比例图为图4.27。



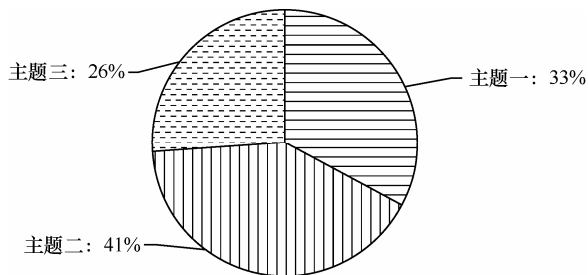


图 4.27 “货币与汇率”专题观点分类比例图

### 主题一：人民币升值导致经济衰退

此分类观点包含三个关键特征词：升值、衰退、经济。对应的观点包括：

观点 1. 人民币升值会使进口商品价格下降，出口商品价格上升，从而刺激进口，抑制出口，减少国际贸易顺差，造成部分企业经营困难，就业减少。

观点 2. 人民币升值使外商直接投资成本上升，对外商直接投资产生一定不利影响，从而制约外商投资。

观点 3. 人民币升值后造成国家外汇储备贬值损失。

### 主题二：人民币升值利益显著

此分类观点包含三个关键特征词：升值、利益、消费。对应的观点包括：

观点 4. 人民币升值提高国内消费者购买能力。

观点 5. 人民币升值有利于国内资本到海外的投资，打破发达国家技术封锁，促进国内企业的技术进步。

观点 6. 人民币升值将增强国外企业投资能力。

主题三：人民币汇率的波动是政府调控的，由大国之间博弈决定，有一定的必然趋势

此分类观点包含三个关键特征词：施压、政府、波动。对应的观点包括：

观点 7. 美国政府施压人民币升值，从中美博弈角度考虑，对中国弊大于利。

观点 8. 人民币汇率波动是由政府控制，由“一篮子货币”所决定。

观点 9. 人民币汇率的制度存在问题，未来会发生改变。

观点 10. 人民币升值是经济发展的必然趋势。

## 2) 观点分类详细分析

为了更好地从宏观向微观的深入分析，本节对每个分类主题内的网络用户分布以及网络民意进行了进一步的分析，希望通过网络用户真实的反馈信息来反映他们对这一些问题的思考与建议。

### 主题一：人民币升值导致经济衰退

在进一步的数据分析与研究中，对网民在这个主题内的深入观点进行梳理与分析，发现存在3个细分的子观点。这些细分的子观点所反映的内容如下：

观点1. 62%的评论认为人民币升值会使进口商品价格下降，出口商品价格上升，从而刺激进口，抑制出口，减少国际贸易顺差，造成部分企业经营困难，就业减少。

例如，“中国作为世界第一大草莓出口国，最近人民币升值这么多，完全出口不出去了。现在市面上草莓泛滥 便宜得很”，“外贸好？哪来的绝世SB，请问单从人民币升值看，成本增加了，为何会更好？一个东西成本3块卖4块，现在成本4块，卖高了人家不要，该高人能不能说明下从那个理论能说明成本增加了生意更好了？更不要说现在外贸有四大恶化点。汇率，地价人工升高，周边第三世界国家同质替代，国外经济不振贸易保护盛行。”

这种基本思想覆盖率较大，对于这种基本的逻辑思维，广大网友还是可以驾驭的。

观点2. 22%的评论认为人民币升值使外商直接投资成本上升，对外商直接投资产生一定不利影响，从而制约外商投资。

例如，“低端产业依赖低工资和低土地价格，但内地工资不停上涨再加上人民币升值，令企业盈利能力进一步降低，外企撤资已经无可避免。”

能有如此深度的网友还是存在的，只不过所占比例较少。

观点3. 16%的评论认为人民币升值后造成国家外汇储备贬值损失。

例如，“2014年第4季度中国外汇储备减少500亿美元！”

如果人民币升值，那么中国巨额外汇储备将会损价。例如，人民币升值10%，从6.6元人民币兑换1美元升至6元兑1美元，中国以人民币计价的外汇储备就减少了10%，这就减少了中国的财富。这种观点从数学上看是正确的，但在经济学上完全错了。中国的财富并不会随人民币的升值而减少。

综上，在本主题中，具有相应子观点的评论分布比例如图4.28所示。

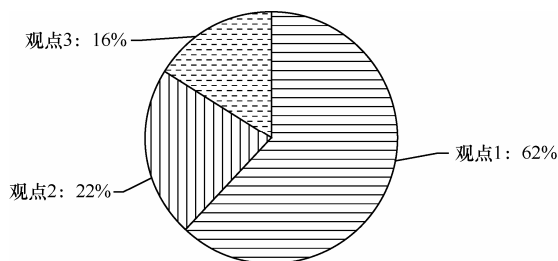


图 4.28 主题一细分子观点比例图

### 主题二：人民币升值利益显著

在进一步的数据分析与研究中，对网民在这个主题内的深入观点进行梳理与分析，发现存在3个细分的子观点。这些细分的子观点所反映的内容如下：

观点1. 55%的评论认为人民币升值提高国内消费者购买能力。

例如，“人民币升值后，中国消费者可以用更少的人民币买到苹果手机？楼主你买过苹果手机吗？没买过也可以去问问价格，别在这里信口胡说好吗？”“近来出境游越来越火热，也有人民币升值的功劳啊，看新闻日本使馆的纸都不够用了。”

大部分网民在BBS中还是只对自己的利益更为关注，他们仅仅能看到的就是人民币升值带来许多国际品牌的“降价”。

观点2. 22%的评论表示人民币升值有利于国内资本到海外的投资，打破发达国家技术封锁，促进国内企业的技术进步。

例如，“人民币升值，有利于产业升级，知识更值钱，有利于白领就业和收入增加升值，不利于低端制造业，不利于蓝领”

有如此见识的网友数量较少，但是分析的还是较为透彻，海外投资带来的必然是产业升级，至于蓝领和白领之分，就仁者见仁智者见智了。

观点3. 23%的评论表示人民币升值将增强国外企业投资能力。

人民币升值后，同样多的人民币对于美元有更大的购买力。大型跨国企业的全球化之路将会得到资金方面的帮助，如华为、联想、海尔等企业。

综上，在本主题中，具有相应子观点的评论分布比例如图4.29所示。

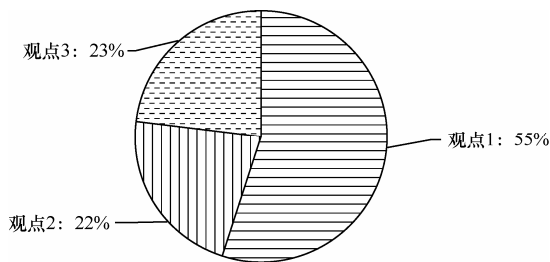


图 4.29 主题二细分子观点比例图

### 主题三：人民币汇率的波动是由政府调控的，由大国之间博弈决定

在进一步的数据分析与研究中，对网民在这个主题内的深入观点进行梳理与分析，发现存在4个细分的子观点。这些细分的子观点所反映的内容如下：

观点1. 36%的评论认为美国政府施压人民币升值，从中美博弈角度考虑，对中国弊大于利。

例如，“你说得对。美国多次施压人民币升值，美国是罪魁祸首。中国以往屈从于美国的压力，一副奴才样。所以这届政府敢于抵住美国的压力，将人民币适当贬值以拯救出口企业，是有魄力的表现。”“好景不长。2004年开始

美国就开始指责中国操纵汇率, 要求人民币升值。你看看, 美国对中国这么多贸易逆差, 说明人民币被低估了, 中国在向美国倾销商品! 没有出口配额, 不需使用关税壁垒, 我照样有办法不让你的商品进来。”

网友分析的还是有一定道理, 美国没有必要做有利于中国政府的决策, 从这一角度考虑, 人民币升值应该是弊大于利。

观点2. 27%的评论认为人民币汇率波动是由政府控制的, 由“一篮子货币”所决定。

例如, “人民币升值, 是ZF人为操纵的, 要么就那破玩意, 屁股纸都嫌硬”。

对于政府的吐槽可能出现在任何地方。即使面对人民币升值的问题, 不仅仅是吐槽, 还有一些戏谑的味道。

观点3. 17%的评论人民币汇率的制度存在问题, 未来会发生改变。

例如, “早就放出风来了, 老美是要我们实行汇率市场化。目前急推的人民币国际化, 就是汇率市场化的前奏。目前的局面搞汇率市场化, 人民币是贬值还是升值, 还有悬念吗?”

对于我们汇率制度, 确实存在诸多问题, 汇率制度灵活性需要提高。

观点4. 20%的评论人民币升值是经济发展的必然趋势。

例如, “美国希望人民币升值是因为出口和顺差, 但是中国如果出现贬值是内因, 是因为热钱外流, 这种经济危机不是一国央行能够干预的, 当年英国央行想干预英镑汇率, 结果在下跌预期下全民换汇, 最后也不得不退出, 任由汇率自由下落, 大趋势一旦确立, 谁挡谁死, 美国央行也不例外。当然, 他也不会干预。”

人民币不升值的后果就是中国的财富源源流向美国, 中国与其他国家的贸易摩擦愈发激烈, 中国的产业结构出现畸形, 减少出口增加消费成为泡影, 国内通货膨胀持续。所以说人民币升值是经济发展的必然趋势。

综上, 在本主题中, 具有相应子观点的评论分布比例如图4.30所示。

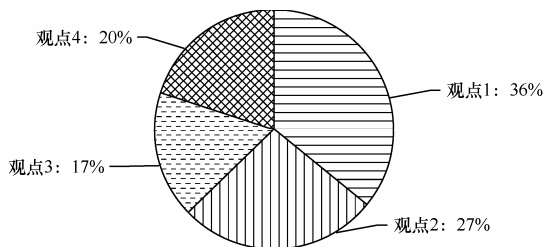


图 4.30 主题三细分子观点比例图

### 3) 观点总结

从网友对于“人民币汇率”问题的讨论中可以看出, 对于汇率的波动, 大

部分网友都有自己的看法，片面的同时也有一定道理。网友更为关心的是汇率波动对于股票、楼市以及自己现金购买力的影响，对于国家层面的相互博弈并没有表现出足够的兴趣。极少部分网友对于汇率问题的分析颇为透彻，同时也得到了较多的回复评论。

汇率的波动事关每个人的利益，在关注汇率波动、分析波动趋势的同时，应该对我们的祖国有信心，相信它可以维护大多数人的利益，灵活调整汇率制度，保证大家的利益不会受到侵害。

### 4.5.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

#### 1) 影响力

专题的传播过程中流量是指相关帖子的回复数量，发展峰值出现在舆论讨论最激烈的时间点。很明显，流量与影响力成正比关系，即流量越大影响力也就越大。因此，本节分别对新增回复数和总回复数按照时间进行统计，然后对统计结果进行可视化得到波线图（图 4.31，图 4.32），以便发现专题传播的发展峰值。

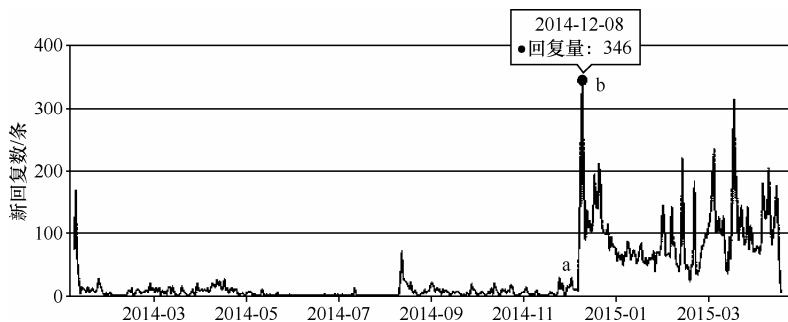


图 4.31 专题日流量时间趋势图

点 a：2014 年 12 月 5 日；点 b：2014 年 12 月 8 日

从图 4.31 可以看出，在点 a 之前，每天对应的新增回复量非常少，说明这一段时间该事件提出来之后并没有引发大的影响；而在点 a 之后，每日新增量开始急剧上升，并且仅仅用了 3 天时间，在点 b 达到了发展峰值，说明该专题在这三天引发了激烈的讨论；在点 b 之后引发的讨论稍微下降，但是偶尔也有上涨趋势，说明该事件从 2014 年 11 月至 2015 年 3 月一直是一个热点事件。在点 a 开始急剧增长，说明该事件由于某个或者多个事件的发生而引发了论坛中激烈的讨论。

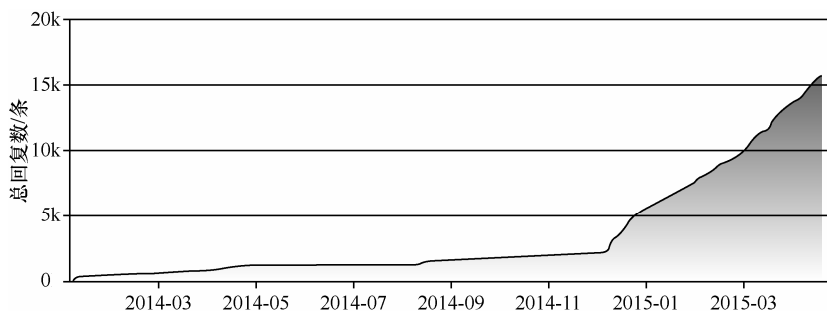


图 4.32 专题总流量时间趋势图

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库，对专题内的回复内容进行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图（图 4.33）和专题情感分类图（图 4.34）。

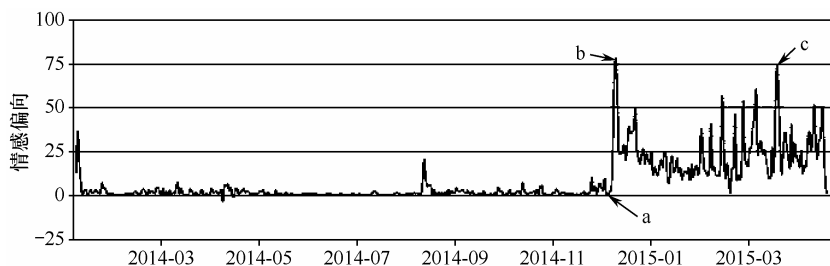


图 4.33 专题情感趋势图

点 a：2014 年 11 月 27 日；点 b：2014 年 11 月 28 日；点 c：2015 年 3 月 23 日

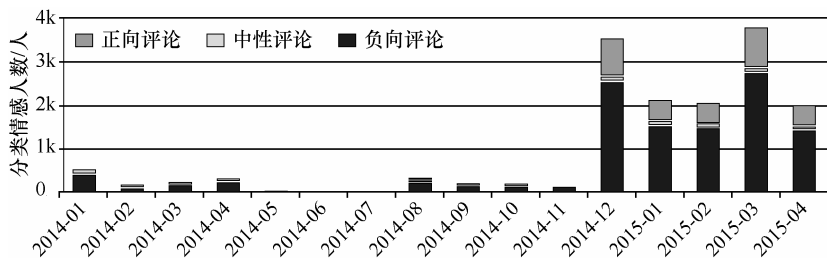


图 4.34 专题情感分类图

从图 4.33 可以看出发展峰值和发展拐点。例如，点 b 和点 c 对应的情感偏向的情感值比较高，说明这两个日达到了情感的发展峰值；而在点 a 之前，情感一直较低，之后发展到峰值，说明这个点属于情感拐点，必定有一些事情影响了

网民的情感。此外，还可以看出在货币领域的网民的情感极易受到影响。

从图 4.34 可以看出，在网民的评论中，负面评论占的比重较大，说明网络环境下，喷子较多，网民发言还是比较随意，大部分网民选择在网络上发泄自己的情绪。

#### 4.5.4 专题用户情感分析

随着近十年来论坛的兴起，网民数量也急剧上升，网民作为网络舆情的一个非常重要因素，行为和言论会引导整个舆论的走向。而在网民中那些发言多、回复多的“意见领袖”或者“领军人物”，往往会主导整个舆论的走向。所以说，探索专题下的人物情感可以帮助我们发现该专题下的情感正向用户，同时也可以帮助论坛找出到处散发消极影响的网民，剔除该用户以净化论坛环境。

本节对各专题下的人物情感进行了统计分析；由于活跃度对情感的影响较大，所以本节构建了分别以情感值和活跃度为坐标轴的专题下的用户散点图（图 4.35）。

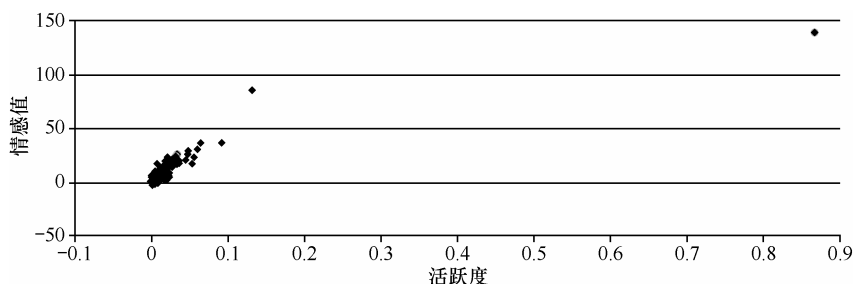


图 4.35 基于活跃度和情感值的用户散点图

圆点代表发帖人，方点代表回帖人

从图 4.35 可以看出，本专题中，发帖人的活跃度相对较低，说明这些发帖人比较少参与话题讨论，只是发表的帖子引起了网民的大量回复，这些发帖人通过自己的热帖影响力很多普通网民，但是并不是真正具有影响力的意见领袖。回帖人的情感值和活跃度分布比较广泛，但绝大部分回帖人都集中在情感值和活跃度较低的区域，这部分人大多数时候只是看到事情的表面，他们的情绪很容易受到所谓“主流声音”的影响，同时也容易受人煽动和利用，也有一些用户的活跃度很高，但情感值比较小，这部分用户有可能是网络中消极言论的传播者，值得注意。

为了进一步进行分析，本节将那些情感值负性较大并且活跃度相对较大的人进行了详细点的描绘，取情感值小于 20，活跃度大于 0.02 的所有用户，得图 4.36。

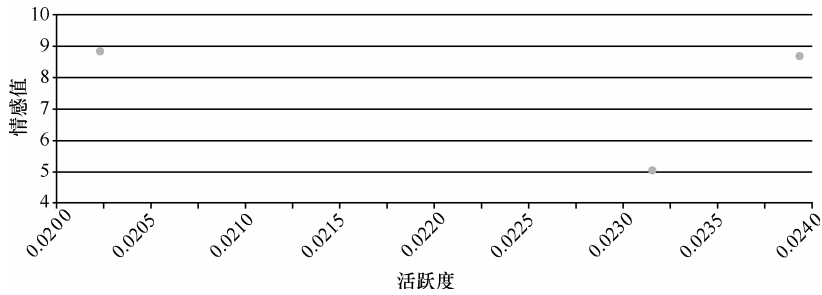


图 4.36 低情感和高活跃度散点图

如图 4.36 所示,有两类点(箭头所指的)比较特殊,第一类点是用户“vspqq”,活跃度比较大,但是他的情感属于最低点,这类用户极有可能是负面情绪的发起者,论坛管理员应该采取相应的操作。第二类点就是图中标出的另外两个点,该类用户活跃度非常高,比其他用户高出很多,但是其情感值却只有 0,这类用户极易被利用,发展成为消极情绪的发起者,应予以关注。

### 4.5.5 专题网络的构建与分析

构建专题网络的方法见节 4.3.5。构建出股市与股票的专题网络,进行可视化,结果见图 4.37。

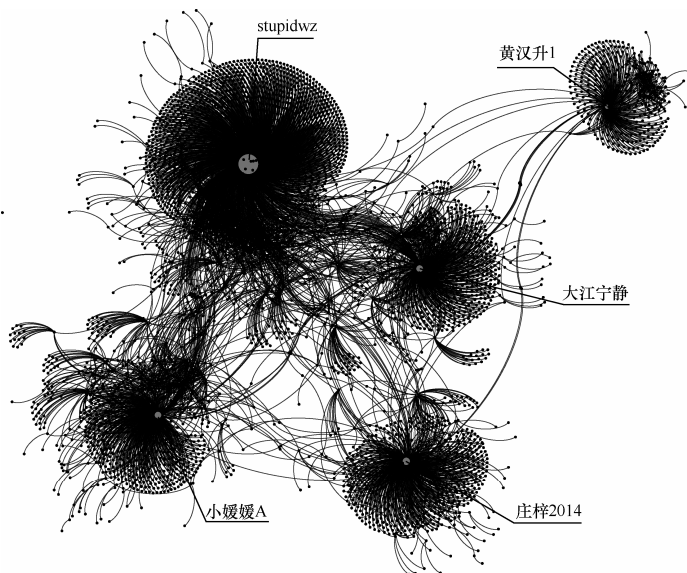


图 4.37 专题传播网络

#### 1) 网络指标概述

本节分别从社会网络的整体性指标(表 4.22)、中心性与影响力指标、关联



系与“小世界”指标进行了分析，以便发现网络中的特殊节点，以便分析个体网络演化与传播。

表 4.22 基本网络指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
3.028	0.002	8	0.733	3.442	0.610

平均度：较小；图密度：较小，该传播网络图比较稀疏；网络直径：较长，该网络的传播较远；模块化指标：划分出 6 个社区，且划分的质量较高；平均路径长度：长度较大，网络传播较广；平均聚类系数：较小。

### 2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算，这里取点度中心度排名前 5 的节点，列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名，得到对应的数据如表 4.23 所示。

表 4.23 中心度指标

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	值排名	绝对值	相对值	值排名	绝对值	相对值	值排名
stupidwz	7587	0.232	1	1466990	9.121	1	2546365	0.158	3957
黄汉升 1	5403	0.165	2	124167	0.772	6	1225609	0.327	4012
小媛媛 A	3089	0.095	3	870612	5.413	3	2548336	0.157	3784
大江宁静	2233	0.068	4	720687	4.481	4	2548208	0.15	3858
庄梓 2014	1597	0.049	5	976849	6.073	2	2548920	0.157	3772

从表 4.23 中可以看出，“stupidwz”的中间中心度远远大于其他节点，排在第一位，说明许多节点都需要经过该节点才能建立联系并获取信息。“庄梓 2014”“大江宁静”“小媛媛 A”的中间中心度也相对比较大，排名分别列在 2~4 位，这三个节点也各自控制了一部分资源，与图 4.37 中的网络图整体趋势一致。“stupidwz”的点度中心度排名第一，与该节点直接相连的节点最多。表 4.23 中列出的 5 个节点接近中心性的值都比较小，说明这几个节点受其他节点的限制或影响很小。

图 4.37 的点度中心势指标为 0.23%，该指标值很小，表示该图在度数上向某点集中的趋势很不明显；该图的中间中心势指标为 9.11%，该指标值较大，表示该图具有中间节点的程度比较大，即在整个网络中有相当部分的节点需要经过某些中间节点，才能得到信息；该图的接近中心势指标为 13.302%，该指标值较大，表示该图的整体联系比较紧密。

### 3) 关联性与“小世界”

图 4.37 的特征途径长度为 2.854，凝聚度为 0.005。该图路径长度的具体分布如表 4.24 所示。

表 4.24 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	6609	0.002	5	417940	0.158
2	407445	0.154	6	79913	0.030
3	466170	0.176	7	14337	0.005
4	1243970	0.471	8	5947	0.002

图 4.37 的聚类系数为 1.301，节点的聚类系数如表 4.25 所示。

表 4.25 聚类系数指标

用户	聚类系数	点对数
stupidwz	0.001	1327635
黄汉升 1	0.003	57291
小媛媛 A	0.004	142845
大江宁静	0.005	147696
庄梓 2014	0.002	178503

4.5.6 个体网络分析

在按照之前的整体网分析之后，我们从中选取那些节点度数相对较大、传播性比较好的帖子，做了某帖子下参与的所有用户的个体网络。本节选取的是“stupidwz”发表的一篇帖子，做可视图（图 4.38），对它进行指标分析，并描述传播路线。

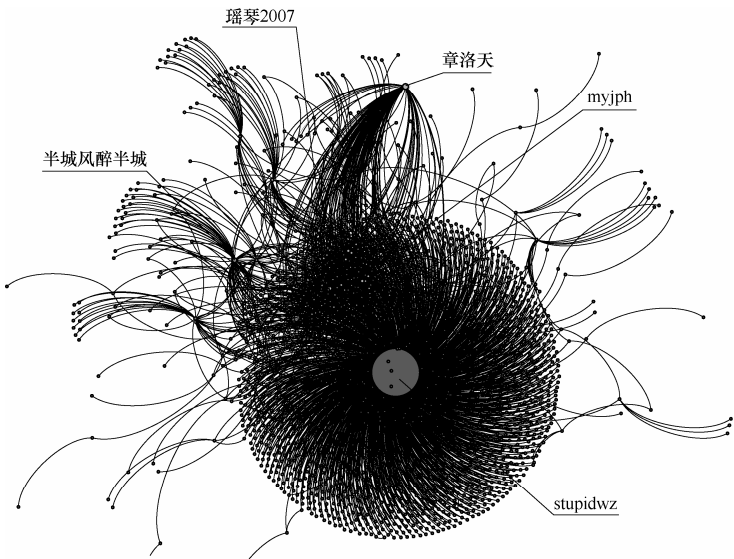


图 4.38 “货币与汇率”个体网络

## 1) 结构洞

这里列举节点度数排名前5位的用户节点的结构洞指标,形成相应的数据指标表(表4.26)。

表 4.26 结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
stupidwz	1631	1469.065	0.901	0.259	0.952
章洛天	111	42.279	0.381	0.787	0.923
半城风醉半城	60	47.926	0.799	0.375	0.729
瑶琴 2007	69	30.700	0.445	0.824	0.876
myjph	45	14.925	0.332	0.895	0.883

从表4.26中可以看出“stupidwz”的节点度数、有效规模和等级度值均为最大,其余用户则远远小于该用户,这说明“stupidwz”居于网络最为核心的位置,而且该节点的效率最高,限制度最小,表明该节点在网络中的行动最高效,受到网络中其他节点的限制比较小。“myiph”对应的效率值较小,限制度较大,说明该节点在网络中受到其他节点的限制比较大,行动力较弱。“章洛天”的等级度比“stupidwz”的值小,又远大于其他节点,位于网络次中心的位置。

## 2) 诚实中间人

这里列出部分代表性用户节点的诚实中间人指标,形成相应的数据指标表如表4.27所示。

表 4.27 诚实中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
stupidwz	100	4950	205	105	32
章洛天	18	153	106	30	17
瑶琴 2007	12	66	45	11	10
yfldll0281x	12	66	39	16	11
半城风醉半城	10	45	36	8	1
ytr567	6	15	8	3	4

HBIO 代表纯中介, HBIO1 代表弱中介, HBIO2 代表非中介。

从表4.27中可以看出“stupidwz”代表的节点共联络了4950对关系,该节点在其中205对关系中作为纯中介,在其中105对关系中作为弱中介,在其中32对关系中作为非中介。“章洛天”“瑶琴 2007”“yfldll0281x”等节点也连接

了数十对节点对，在其中充当了中介。可以看出，“荣耀与卑微”发表了帖子后，大量用户参与的讨论。在回复的用户中，“章洛天”“瑶琴 2007”和“yfldll0281x”等用户发表的评论引起了部分网民的局部讨论，产生了一定的影响力。

## 4.6 本章小结

本章主要针对天涯论坛的经济领域的舆情进行了分析，选取了网民关注最广泛的房市与房价、股市与股票和货币与汇率三个专题，分别从内容和结构上进行了深入的研究和分析。

以经济建设为中心，是党基本路线的基石，是我国经济社会发展的一条主线。我们要紧紧围绕使市场在资源配置中起决定性作用深化经济体制改革，坚持和完善基本经济制度，加快完善现代市场体系、宏观调控体系、开放型经济体系，加快转变经济发展方式，加快建设创新型国家，推动经济更有效率、更加公平、更可持续发展。

通过针对网络中的舆情分析，我们发现虽然近年来我国经济持续高速发展，但在发展过程中仍然存在一些问题，如何做到经济又好又快地展，是值得各界人员深入思考的问题。通过网络的舆情分析也发现，相关的领域的问题也成为了网民关注的焦点领域，一些网民不仅只对目前的问题进行了深入的思考与挖掘，而且还有一些网民甚至对一些问题提出了理性的解决方案，因此，在网络时代，如何让更多的民众通过网络等不同的途径进行更加有效地、理性地参政和议政，这也是本报告进行研究的一个方向。

## 第 5 章 政治领域的相关舆情分析

所谓政治体制改革，就是在社会主义政治总格局和权力结构形式不变的前提下，调整部分社会关系，提高政治与行政管理工作的效能，对政权组织、政治组织的相互关系及其运行机制的调整和完善，从而促进社会经济、文化等事业的综合发展。其目标包括以下三个方面：第一，巩固社会主义制度；第二，发展社会主义社会的生产力；第三，发扬社会主义民主，调动广大人民的积极性。随着我国政治体制改革的不断深入，我国政治领域发生了显著变化，在建设社会民主政治的道路上迈出了坚实的步伐。但是，相关的组织制度以及工作制度等领域已出现一些与当前社会与经济发展不匹配等问题。因此，需要通过不断地调整与优化，将经济基础与上层建筑之间存在的一些不合理与不匹配进行优化和改革。

目前政治领域的改革思路可以概括为一个准则、三个目标和十项任务。一个准则是指坚持党的领导、人民当家做主、依法治国为政治体制改革的行为准则。发展社会主义民主政治、建设社会主义法治国家、强化权力运行制约和监督体系视为政治体制改革的三个主要目标。推动人民代表大会制度与时俱进、推进协商民主广泛多层制度化发展、发展基层民主视为发展社会主义民主政治的三项主要任务，将维护宪法法律权威、深化行政执法体制改革、确保依法独立公正行使审判权检察权、健全司法权力运行机制、完善人权司法保障制度视为建设社会主义法治国家的四项主要任务，将形成科学有效的权力制约和协调机制、加强反腐败体制机制创新和制度保障、健全改进作风常态化制度视为强化权力运行制约和监督体系的三项主要任务。

为了客观地分析政治领域的热点舆情，我们从天涯论坛中选取了 2014 年度一些政治领域热点专题的帖子进行深入的总结和分析，本章后续部分将从不同的角度展示该领域舆情分析的结果。

### 5.1 政治领域热点专题热点词分析

本节选取政治领域的相关帖子进行分析，通过对词频进行统计得到该领域的热点词及演化情况，通过对原帖进行文本聚类得到热点事件，从而确定该领域的热点专题。

### 5.1.1 政治领域热点词分析

热点词出现的频率会随着热点事件的发生而快速增长,从热点词的演化过程可以看出不同时间热点事件的主题变化情况。对2014年全年的数据按季度进行词频统计和排名,结果如表5.1所示。

表 5.1 高频词汇统计

2014.01~2014.03		2014.04~2014.06		2014.07~2014.09		2014.10~2014.12	
词语	词频	词语	词频	词语	词频	词语	词频
北京	95948	北京	89045	北京	101700	北京	176007
中国	45249	国家	34994	中国	48908	中国	150454
国家	24552	首都	32990	首都	26700	国家	62556
首都	23160	中国	28978	国家	25825	首都	60587
经济	19593	城市	21216	经济	20649	经济	48774
人口	18289	人口	20149	人口	19365	城市	39343
城市	15400	经济	19563	疏解	18500	人口	35698
市场	13719	产业	18965	城市	16267	机构	34459
政府	13702	市场	18991	机构	15192	产业	39170
全国	12594	政府	18483	市场	14602	市场	34689
天津	11135	全国	17283	时间	13400	举报	28902
长江	11079	问题	15737	政府	12845	问题	27087
建设	10595	战略	15636	举报	12818	天津	26614
官员	9632	举报	15326	总部	11762	长江	25587
中央	9532	天津	15245	决定	11042	中心	25387
企业	9433	长江	15004	作者	10782	中央	24741
利益	8742	建设	14720	回复	10749	官员	24476
承接	8055	中心	14289	企业	10670	利益	22103
核心	7245	官员	13567	中央	10038	襄阳	20456
规划	7149	企业	12956	选择	8757	洛阳	19162
阶层	7149	承接	11205	核心	7632	承接	19057
转移	7119	核心	9999	空间	6465	污染	13100
北京市	7072	阶层	9863	南阳	6402	政治	11904
迁移	7048	转移	9853	达到	5596	文化	10987
空间	6117	迁移	9800	污染	5380	民族	10734

续表

2014.01~2014.03		2014.04~2014.06		2014.07~2014.09		2014.10~2014.12	
词语	词频	词语	词频	词语	词频	词语	词频
行政	5319	工程	8845	政治	4863	机关	10425
控制	5208	工业	8739	改革	4830	改革	9845
搬迁	5039	缺水	8395	民族	4465	重庆	9752
改革	5009	水	7161	交通	4381	北方	9157
世界	4598	权贵	7099	河北	4362	南京	9247
历史	4457	世纪	7036	区域	4310	科教文卫	9084
民族	4305	污染	6459	南京	4291	俄罗斯	8238
文化	4289	改革	6206	科教文卫	4268	京津冀	7791

为了更好地观察高频词的演化过程，我们用作图工具 NetDraw 对高频词和时间的对应关系进行可视化，得到了高频词演化图（图 5.1）。

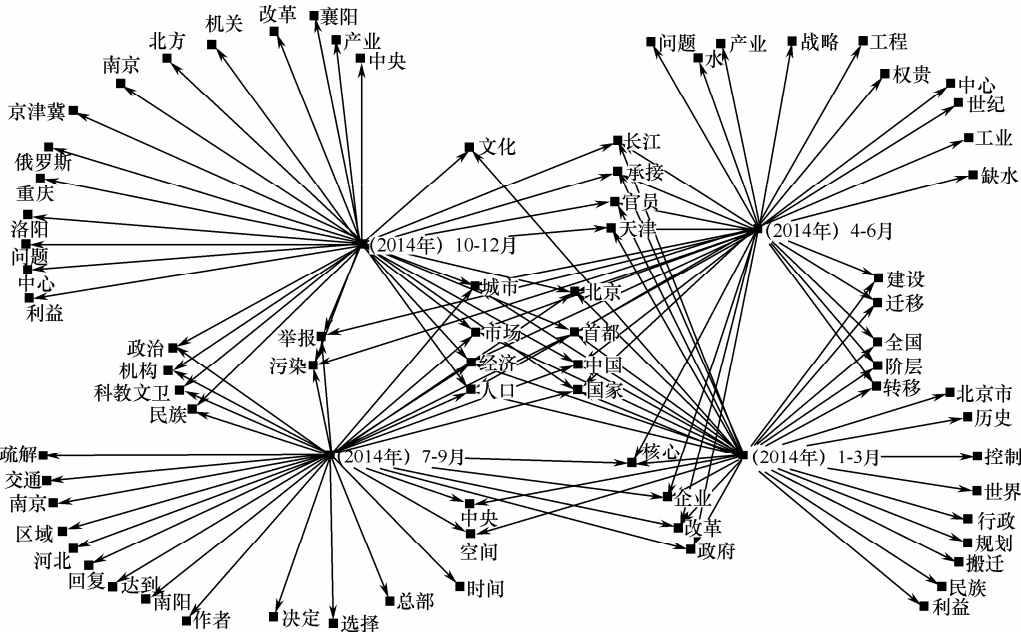


图 5.1 高频词演化图

从图 5.1 中可以看出，“北京”“国家”“首都”“人口”等词语为 2014 年的年度热词，即这些词在四个季度中都作为高频词出现，说明这些词语所涉及的领域和话题在 2014 年中有着较大的影响和热度。“南阳”“河北”“洛阳”等词汇出现在不同季度的热词中，说明在特定时间、特定地点发生的一些政治事件

成为网民们讨论的热点。“改革”作为年度热词与本章的“国企改革”专题有着较高的结合度。“举报”在第二季度成为季度热词后一直持续至年末，与本章的“反腐风暴”专题具有较高的结合度。

5.1.2 政治领域热点专题分析

通过对 2014 年政治领域的帖子进行文本聚类得出事件，根据 2.3.2 节的指标进行事件热度的计算，并对这些事件进行摘要的抽取和标题的生成。本节列举热度排名前 10 点事件形成表 5.2。

表 5.2 2014 年全年政治领域十大热点事件

热点事件	帖子量/条	总浏览量/条	总回复量/条	参与人数/个	专题热度
×××浮出水面，反腐大幕刚刚拉开	2	1043654	7201	3698	0.999990
反腐将打造出一个怎样的新中国？	4	506874	2458	1210	0.813254
四中全会依法治国问题前瞻	2	104896	1982	698	0.759304
“依宪治国”与“依法治国”	2	275108	1485	759	0.739344
没有反腐制度跟进的打虎，究竟还能走多远？	3	125478	1047	658	0.532582
当年国企改革错误的焦点不在于工人下岗	2	102306	341	249	0.410269
“小官巨贪”都不能治还谈什么“依法治国”!!!	1	275698	678	514	0.501238
企业女职工在何岗位及何时退休应由谁说了算？	2	225417	568	354	0.463288
养老金并轨的关键词“赎买”	3	205412	478	396	0.459850
国际追逃风暴与依法治国	2	146325	199	107	0.432640

从表 5.2 中可以发现，“×××案浮出水面，反腐大幕刚刚拉开”“反腐将打造出一个怎样的新中国？”和“没有反腐制度跟进的打虎，究竟还能走多远？”等事件集中反映了反腐问题；“四中全会依法治国问题前瞻”“‘依宪治国’与‘依法治国’”和“国际追逃风暴与依法治国”等事件集中反映了依法治国问题；“当年国企改革错误的焦点不在于工人下岗”“企业女职工在何岗位及何时退休应由谁说了算？”等事件直接或间接地反映了国企改革中的问



题。综上所述，国企改革、反腐问题和依法治国是政治领域中网民讨论和关注较多的几个热点专题，本报告以国企改革、反腐问题和依法治国三个专题为例，对政治领域的网络舆情进行深入的分析。

5.2 政治领域意见领袖分析

在政治领域中涉及的用户包括相关帖子的发布者以及参与浏览和回复这些帖子的舆情受众。在信息传播中，一些网民凭借自身表达能力突出、见解深刻等素质，成为政治领域中的“意见领袖”。他们善于从表象信息透析到比较深层的价值内涵和意义，并能对问题进行解析，形成自己的观点并加以表达，他们的观点能够引起众多普通网民的兴趣和参与，促成网络舆论的最终形成。

通过对网络舆情发布者的关注度分析，可以有效地监测出网络舆情在传播与演化过程中具有影响力的意见领袖。本节采用发布者影响力、发布者活跃度和发布者价值这三个不同维度的指标对舆情发布者进行分析，采用负面指数、参与度和受众影响力对各个专题的舆情受众进行分析。

5.2.1 意见领袖

1) 发布者影响力

根据节 2.3.2 中提到的计算公式，对网络用户的影响力进行计算，其结果如表 5.3 所示。

表 5.3 政治领域论坛用户影响力排名

用户	粉丝数/个	帖子点击数/次	发帖数/条	帖子回复数/条	回帖数/条	影响力
野渡自渡人	6841	1526981	35	10328	1724	0.935
cherry521314	6	3160674	2	949	2	0.876
说真话好难	519764	39533	4	124	7	0.859
jzqtj123	22496	408758	24	4572	1182	0.647
cat_alx	17	290992	1	4669	191	0.589
天涯海角客 1	4126	1209032	12	1229	0	0.582
新手 L	6	626817	1	3149	13	0.564
OK 千里千寻	3	825089	1	1734	1	0.504
鸭绿江上的渔夫	18	923955	4	1241	111	0.492
冷库的叫兽	11	762501	2	1637	2	0.476

续表

用户	粉丝数/次	帖子点击数/条	发帖数/条	帖子回复数/条	回帖数/条	影响力
3bbdad6a34334d8891ea391cdb66c84f	25441	419 981	12	2238	1	0.471
朴坚实	5	1043628	1	392	0	0.453
f51f1a86edeb406eac6250974027cc57	0	739706	1	1345	0	0.441
df2bc6584c634d669fa98d96209ae4a0	4081	139923	9	3308	13	0.436
金健米业总经销	8	1036241	1	0	0	0.413

表 5.3 列出了影响力前 15 位的用户。“野渡自渡人”的影响力最高，其发帖数、帖子回复数和回帖数也最多，可将其视作意见领袖之一；“cherry521314”和“说真话好难”分别由于浏览数和粉丝数较大排名位于第二位和第三位。从此表中还可以看出，浏览数和粉丝数数值过大，对于影响力影响较大，此种情况不太合理，计算公式还需要进一步改进。

2) 发布者活跃度

公式具体解释见节 2.3.2 节。这里， $g_1$  取 0.6， $g_2$  取 0.4，计算结果如表 5.4 所示。

表 5.4 政治领域论坛用户活跃度排名

用户	发帖数/条	回帖数/条	活跃度
东方夜话 2014	37	1867	1.0000000000000000
野渡自渡人	35	1724	0.924057217165150
济南人 2014	2	1549	0.807282184655397
徐晖 2014	1	1343	0.699349804941483
jzqtj123	24	1182	0.633550065019506
南乡 2014	35	1040	0.568270481144343
光明客上访	40	998	0.550325097529259
大路朝天亮	1	977	0.508972691807542
13627711818	1	887	0.462158647594278
zushenliang	19	763	0.411703511053316
易天校园 2011	3	764	0.399739921976593
别了还来	30	646	0.359427828348505
陕北黑脸汉子	3	548	0.287386215864759
hebeilt	1	502	0.261898569570871
挡不住的陈文军	12	458	0.247594278283485

表 5.4 列出了发布者活跃度前 15 名的用户，表明这些用户在论坛中较为活

跃，同时可以作为重点监视对象。“野渡自渡人”属于活跃性的意见领袖，保持较高的发帖量和回帖量，同时具有较高的影响力。用户影响力和用户活跃度并不存在严格的正比关系，影响力高的用户活跃度不一定高。

3) 发布者价值

网络中发布者之间会经常互动，发布者的情感价值观也会相互传递、相互影响。发布者的情感倾向成为舆情监测的重点。

选择排名前 10 的负向情绪用户（表 5.5），可以看出大多数情感倾向为负值的用户，已经被网站拉黑或者清除，没有被拉黑的用户也可以作为舆情检测的重点对象。

表 5.5 作者情感倾向

发布者	情感倾向	发布者	情感倾向
匿名用户 201	-4.66667	匿名用户 409	-2.67500
匿名用户 535	-4.00000	匿名用户 737	-2.40000
匿名用户 643	-3.33333	rollaswold	-1.95000
inkeee	-3.03750	土老巴 2011	-1.80000
匿名用户 250	-2.80000	stupid_fox	-1.67867

注：爬取数据时不能获取用户名的用户被标注为匿名用户。

5.2.2 舆情受众

舆情受众主要指所有被舆情影响到的网络用户。衡量舆情受众主要有三个指标：负面指数、参与度、受众影响力。（指标的计算公式同节 4.2.2）

本章热点专题的舆情受众指标如表 5.6 所示。

表 5.6 专题舆情受众表

热点专题	负面指数	参与频度	受众影响力
国企改革	0.5214	0.001324	0.069872
反腐风暴	0.4985	0.002487	0.046201
依法治国	0.5471	0.004017	0.098120

5.3 专题 1：国企改革

经过 5.1 节的热点专题与热点词分析，选取国企改革作为政治领域的第一个热

点专题。国企的改革和发展，对于建立社会主义市场经济体制，促进经济持续快速健康发展，提高人民生活水平，保持安定团结的政治局面，巩固社会主义制度，都具有十分重要的意义。国企改革作为 2014 年的政治热点事件之一，取得了许多成就，但是要完成制度创新和改革的纵深推进还需要党和政府更多的指导，需要人民更多的监督。本节将从国企改革专题的发展简介、专题中网民观点的分类和总结、专题的传播趋势和用户情感以及专题网和个体网分析等方面进行详细叙述。

5.3.1 专题发展简介

2014 年对于中国是特殊之年，经济回落，新常态渐成共识，而国资央企仍表现抢眼，亮点频现。

中央、地方都在为国企改革行动。国务院牵头成立了国企改革领导小组，基本完成了国企改革方案的顶层设计。多家中央企业“涉水过河”，超过 25 个省市各显神通，发力“混改”。混合所有制改革正式起步，董事会放权正在试点，薪酬改革破冰伊始，国有资本投资公司任重道远。2014 年，国企重拳反腐与法治建设并行，规范、透明成为国有企业的全新标签。广大网民对于国企的“垄断性”颇有不满，但对于国企改革抱有很大期待。

表 5.7 国企改革专题热帖

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
越南排华河静中冶牺牲的工友证明，危局面前只有国企才有战斗力	huazhong295	2014/05/21	110931	959	284	180	99.228
【逆行斋语】改良派、革命派与中国之命运	fae79a84d7774c6e-813c6ea5853515e3	2014/08/07	54477	667	121	124	33.369
2014 年，中国人挺起了胸膛	lixin330a	2014/12/29	29632	451	198	133	80.010
骂总理的人，基本都是这个社会的吸血虫	船老大的爱国心	2014/04/12	17744	454	103	108	13.167
乌克兰，乌克兰，乌克兰——一个失败的典范	顿河骑兵	2014/02/20	164619	406	212	73	4.074

### 5.3.2 观点分类与观点列表

以论坛中网民的回复为研究对象,通过观点挖掘和语义分析形成本专题内的核心观点,并对其进行分类和详细描述。

#### 1) 观点分类

根据观点涉及的特征词汇、领域以及网民发表的各个观点内容本身之间存在的相关性与差异性,可以将所有评论分为以下四个主题。

##### **主题一: 国企改革势在必行, 支持国企改革及国家层面的相关政策**

此分类观点包含三个关键特征词: 国企、改革、政策。对应的观点包括:

观点 1. 国企改革是促进经济持续快速发展、提高人民生活水平的必要措施。

观点 2. 国企是国民经济的支柱, 它们的良好发展利国利民。

##### **主题二: 国企改革事关重大, 应该妥善处理, 谨慎处置**

此分类观点包含三个关键特征词: 妥善、谨慎、改革。对应的观点包括:

观点 3. 国企改革事关重大, 会影响国家未来的发展。

观点 4. 国企改革是好事, 但是一定要防止国有资产流失。

观点 5. 国家领导人对于国企改革足够重视, 借鉴他国的经验教训我们才能成功。

##### **主题三: 国企改革导致国有资产流失, 造成少数人的富裕, 不支持**

此分类观点包含三个关键特征词: 流失、国有资产、不支持。对应的观点包括:

观点 6. 国企改革的结果就是少数人“侵吞”国有资产, “中饱私囊”。

观点 7. 当年国企改革的结果是工人的失业与国有资产的流失, 现在的改革很有可能重蹈覆辙。

观点 8. 宁愿不进行国企改革, 也不愿国有资产的流失。

##### **主题四: 国企改革是国家层面的事务, 与普通老百姓没有关系**

此分类观点包含四个关键特征词: 装修、风格、选择、问题。对应的观点包括:

观点 9. 国企改革与否都是国家层面的事务, 与老百姓关系不大, 利益相关性也不大。

观点 10. 自己的言行对于国企改革没有影响, 但是依然不影响对于国企改革

革的关注和信心。

从图 5.2 可以看出,大多数网民支持国企改革(主题一占比 35%,主题二占比 36%),对国企改革的相关政策也非常关心,具有“忧国忧民”的良好品质。同时有一些网民对于国企改革提出了自己的看法和意见,依据上一次国企改革对工人阶级造成的生活震荡(主题三占比 18%),他们有权利提出疑问和质疑。还有一少部分网民认为国家大事是政府和党的事情,和自己无关,此种“小民思想”所占比例不多(主题四占比 11%)。

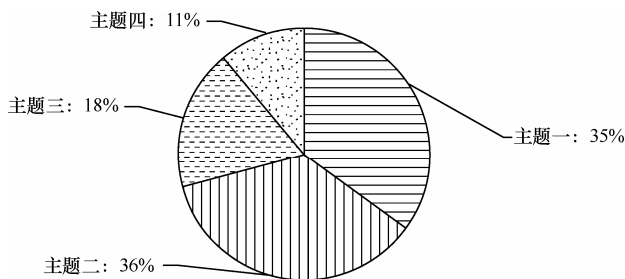


图 5.2 “国企改革”专题观点分类比例图

## 2) 观点分类详细分析

为了更好地从宏观向微观的深入分析,本节对每个分类观点内的网络用户分布以及网络民意进行了深入分析,希望通过网络用户真实的反馈信息来反映他们对这一些问题的思考与建议。

### 主题一: 国企改革势在必行, 支持国企改革及国家层面的相关政策

在进一步的数据分析与研究中,对网民在这个主题内的观点深入梳理与分析,发现存在两个细分的子观点。这些细分的子观点所反映的内容如下:

观点 1. 32%的评论认为国企改革是促进经济持续快速发展、提高人民生活水平的必要措施。

例如,“自以为是,中国人绝对不答应,坚定支持国企改革。”“坚定支持国企改革。”

这样的选择结果具有其存在的理由,目前许多垄断性质的国企积弊难返,如果不进行国企改革,会导致国企的综合竞争力进一步下降,有朝一日外企和私企进入该领域后,后果不堪设想。

观点 2. 68%的评论认为国企是国民经济的支柱,它们的良好发展利国利民。

例如，“国企和民企不是对立的关系，关注国企改革，他们能干的往往是一些大项目，一投就几十亿元、上百亿元，一般私企干不了。”

国企是国民经济的基础，国家财政收入的纳税大户，应该通过改革使得国企更有活力，更好服务社会、服务人民。

综上，在本主题中，具有相应子观点的评论分布比例如图 5.3 所示。

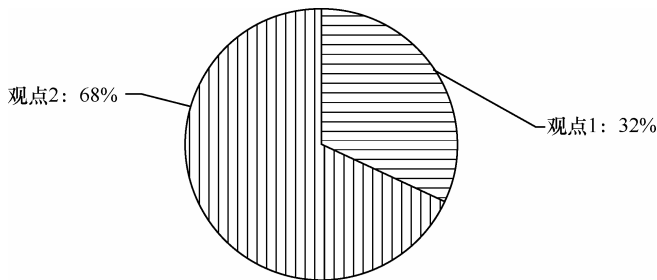


图 5.3 主题一细分子观点比例图

## 主题二：国企改革事关重大，应该妥善处置，谨慎处置

在进一步的数据分析与研究中，对网民在这个主题内的观点深入梳理与分析，发现存在 4 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 40%的评论认为国企改革事关重大，会影响国家未来的发展。

例如，“说句老实话，国企改革稍微出点差池，中国就有可能落得和原苏联一个下场，望上面的人且重且珍惜，莫为一己私利，害了整个国家。”

部分网民的政治觉悟还是有的，关心国家大事，忧国忧民。

观点 2. 28%的评论表示国企改革是好事，但是一定要防止国有资产流失的现象。

例如，“如果不防这一点，乌克兰、曾经的俄罗斯寡头也会在中国出现。国企改革不是个简单的事，老大有眼光，你也有眼光。”

网民从乌克兰和俄罗斯的例子中发现了“寡头”出现的原因，提醒党和政府防止这种情况的发生。

观点 3. 32%的评论表示国家领导人对于国企改革的足够重视，借鉴他国经验教训我们才能成功。

例如，“所以老大说了，这次‘要吸取过去国企改革经验和教训，不能在一片改革声浪中把国有资产变成牟取暴利的机会。改革关键是公开透明。’”

这些关注时事同时又保持乐观的网民们对于国家相当有信心！满满正能量，赞一个！

综上，在本观点中，具有相应子观点的评论分布比例如图 5.4 所示。

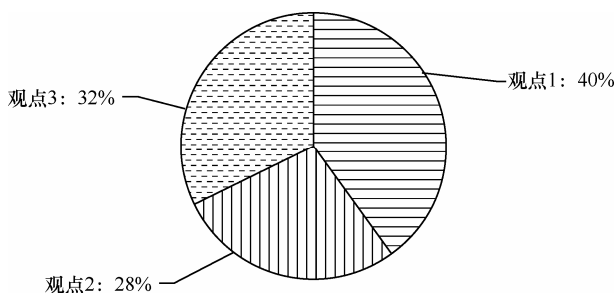


图 5.4 主题二细分子观点比例图

### 主题三：国企改革导致国有资产流失，造成少数人的富裕，不支持

在进一步的数据分析与研究中，对网民在这个主题内的观点深入梳理与分析，发现存在 3 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 46%的评论认为国企改革的结果就是少数人“侵吞”国有资产，“中饱私囊”。

例如，“要的就是国企改革失败，公有财产大量流失被少数人侵吞暴富，已经讲了一部分人富起来。”

网友对于国有资产的流失深恶痛绝，有一种“割肉”般的感觉。他们愤怒地认为国企改革就是当权者“中饱私囊的游戏”。

观点 2. 16%的评论认为当年国企改革的结果是工人的失业与国有资产的流失，现在的改革很有可能重蹈覆辙。

例如，“当年的国企改革是人民心中永远的痛！一辈子不忘！”

用当年改革的结果来决定现在是否改革似乎不太合理，但是当年的教训是党和政府应该借鉴的。

观点 3. 38%的评论认为宁愿不进行国企改革，也不愿国有资产的流失。

例如，“顶！现在国企改革不是要将国企变私企，如果变私企、外企我还是选择宁可不改革！”

部分网友还是比较悲观的，其实国企改革是一条可行性更强的道路。

综上，在本主题中，具有相应子观点的评论分布比例如图 5.5 所示。



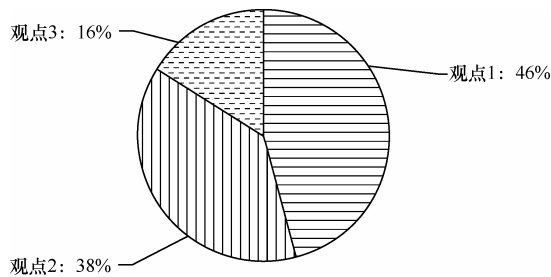


图 5.5 主题三细分子观点比例图

**主题四：国企改革是国家层面的事务，与普通老百姓没有关系**

在进一步的数据分析与研究中，对网民在这个主题内的观点深入梳理与分析，发现存在 2 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 55% 的评论认为国企改革与否都是国家层面的事务，与老百姓关系不大，利益相关性也不大。

例如，“其实国企，怎么说呢，国企明着说是全体人民的，但盈利都不顾及人民。为什么国企改革，人民反响不强烈。国企都被权贵家族霸占着，成为家族企业。名为国企实为私企。反正无论怎么折腾，那是老虎打架，我们无能为力，不管怎么改，与屁民没有多大关系……”

这部分网民看似“合理”的想法其实充满了负能量。“国家兴亡，匹夫有责”，国企改革这种关乎民生社稷的大事都不能积极主动地去参与，国家有危难的时候还能指望你吗？

观点 2. 45% 的评论认为自己的言行对于国企改革没有影响，但是依然不影响对于国企改革的关注和信心。

例如，“我们说话当然不管事，但是老大说话应该管事吧。你看看他关于国企改革的发言就明白了。”

虽然没有“影响它的能力”，但阻挡不了我关注它的想法！正能量！不过你对它的影响可能比你想象中的要大一点。

综上，在本观点中，具有相应子观点的评论分布比例如图 5.6 所示。

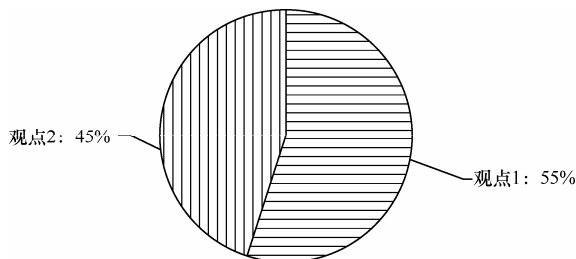


图 5.6 主题四细分子观点比例图

### 3) 观点总结

国企改革作为十八大后的“高频词”反映出党和政府对于深化国有企业改革、完善各类国有资产管理体制的决心和信心。广大网民对于国企改革也是意见纷纷。大部分网民对于国企改革持有支持和乐观的态度，同时有网民对于国企改革提出了相应的质疑，他们认为国企改革期间国有资产的流失太多，政府的监管力度太差。国企改革已经是大势所趋，是不断增强国有经济活力、控制力、影响力的必要措施，希望在新一轮的国企改革中加大监管力度，阻止国有资产流失的同时使国有企业改革取得成功。

#### 5.3.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

##### 1) 影响力

专题的传播过程中流量是指相关帖子的回复数量，发展峰值出现在舆论讨论最激烈的时间点。很明显，流量与影响力成正比关系，即流量越大影响力也就越大。因此，本节分别对新增回复数和总回复数按照时间进行统计，然后对统计结果进行可视化得到波线图（图 5.7，图 5.8），以便发现专题传播的发展峰值。

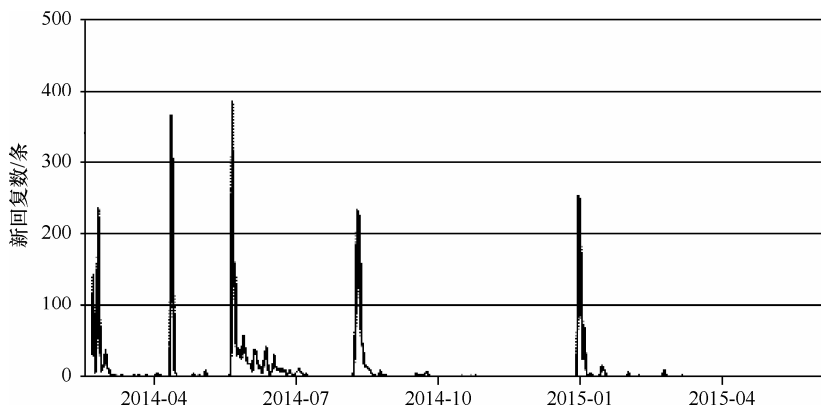


图 5.7 专题日流量时间趋势

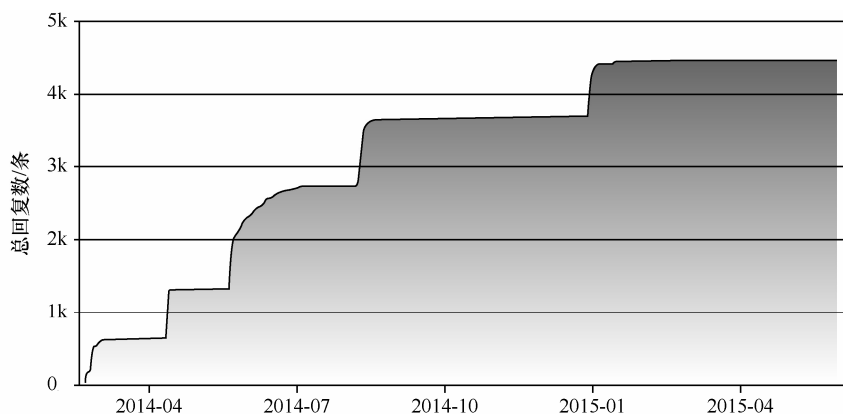


图 5.8 专题总流量时间趋势图

从图 5.7、图 5.8 中可以看出，此专题的流量集中在 2014 年 2 月 24 日、4 月 12 日、5 月 21 日、8 月 10 日、12 月 30 日。在专题日流量中，这几天的新回复数量较大；在专题总流量中这几天回复数的增速较大。国企改革专题在 5 月 21 日的日流量达到最大值，回复值为 388；5 月 21 日过后，该专题的总流量在数十天之内依然保持增长，说明 5 月 21 日广大网友交流较多，该天的话题同时引起了后续的讨论，导致该专题的总流量持续增加。

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库，对专题内的所有回复内容进行情感分析。情感分析包括对专题内情感总体偏向（图 5.9）的分析和专题内回复内容的情感分类（图 5.10）的分析。

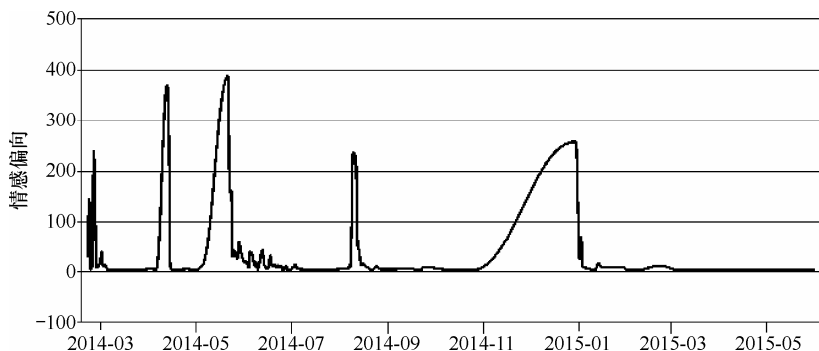


图 5.9 专题情感趋势图

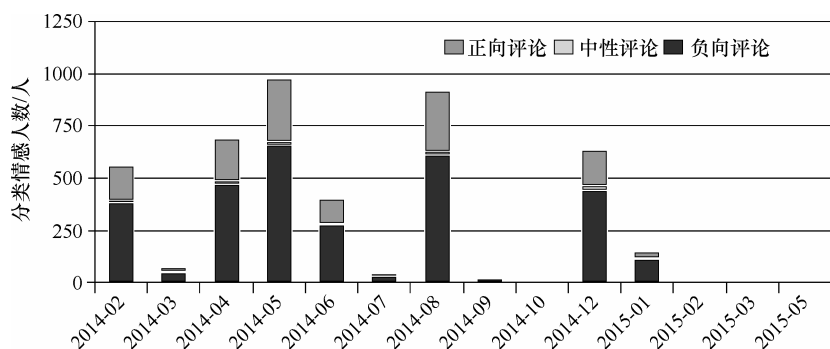


图 5.10 专题情感分类图

从图 5.9 可以看出发展峰值，2014 年 2 月 24 日、2014 年 4 月 12 日、2014 年 5 月 21 日、2014 年 8 月 10 日、2014 年 12 月 30 日事件情感值较大，同样这几天所对应的日流量数值较大。因为较大的日流量才会使当日的情感偏向波动较为明显，较小的日流量会情感偏向趋向于零。

从图 5.10 中可以看出，负向评论的数量远大于正向评论和中性评论，这是由于网络环境与语言的随意性和较低约束性造成的。

### 5.3.4 专题用户情感的分析

随着近十年来论坛的兴起，网民数量也急剧上升，网民作为网络舆情的一个非常重要因素，行为和言论会引导整个舆论的走向。而在网民中那些发言多、回复多的“意见领袖”或者“领军人物”，往往会主导整个舆论的走向。所以说，探索专题下的人物情感可以帮助我们发现该专题下的情感正向用户，同时也可以帮助论坛找出到处散发消极影响的网民，剔除该用户以净化论坛环境。

本节对各个专题下的人物情感进行了统计分析，由于活跃度对情感的影响较大，所以本节构建了分别以情感值和活跃度为两个坐标轴的专题下任务的散点图（图 5.11）。

从图 5.11 可以看出，一般发帖人的情感值较高，那些活跃度较高同时情感值较高的用户可以被视作意见领袖。同时存在一些不是发帖者但其活跃度较高的点，可能是该用户的回复内容引起了其他网民的讨论，也可能是该用户经常在论坛中灌水、评论，具有较高的回复参与量，但是通常回复的内容情感性较低，感情色彩不明显。

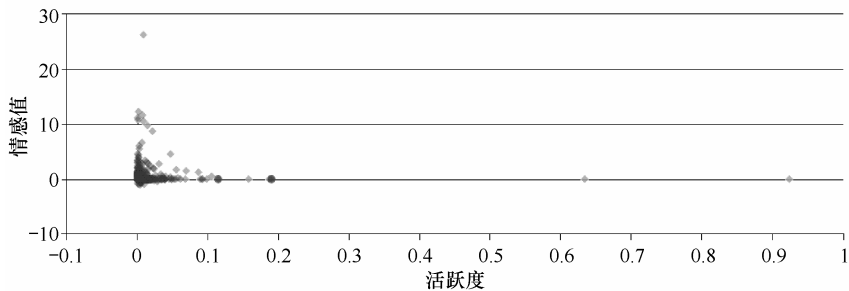


图 5.11 基于活跃度和情感值的用户散点图

圆点代表发帖人，方点代表评论人

为了进一步进行分析，本节将那些情感值负性较大且活跃度较大的用户进行了详细点的描绘，取情感值小于 20，活跃度大于 0.02 的所有用户，作图 5.12。这是一类在论坛中活动积极、负面情绪较大的用户。这类用户的帖子和评论影响到其他人对国企改革的想法。

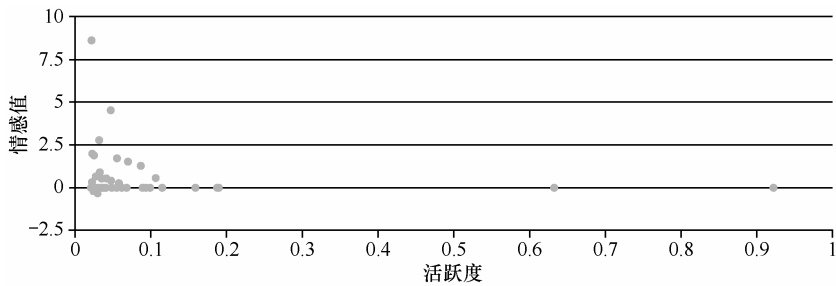


图 5.12 低情感和高活跃度散点图

从图 5.12 可以观察到，那些活跃度较高同时情感值较低的用户。论坛管理员要对这些用户进行监察和历史查询，当这些用户出现不正当言论时进行及时的处理。

5.3.5 专题网络的构建与分析

构建专题网络的方法见节 4.3.5。

国企改革专题构建的专题网络结果见图 5.13。

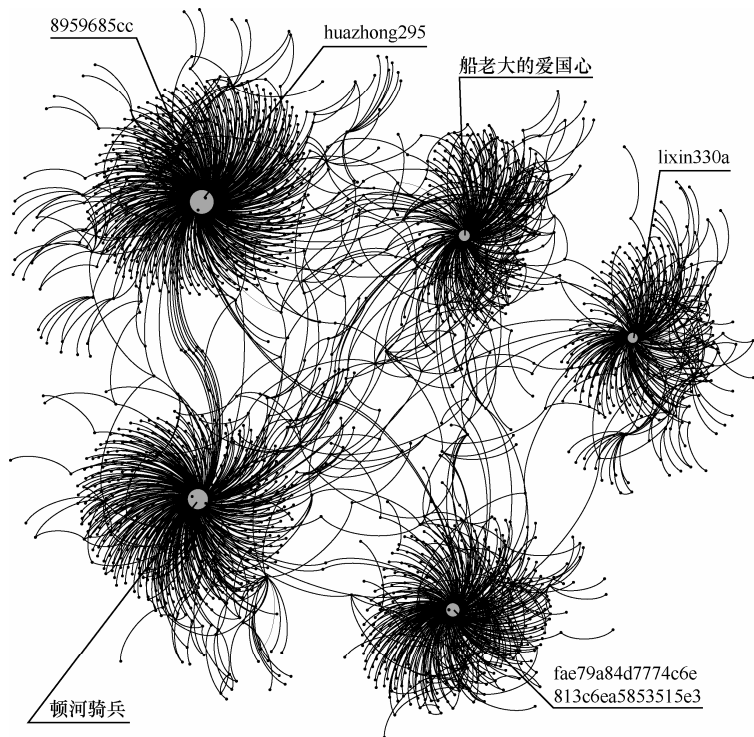


图 5.13 专题传播网络

由图 5.13 可知，该主题下有五个较大的社区，选取“顿河骑兵”和“huazhong295”发布的帖子及其受众所构成的社区进行分析。

1) 网络指标概述

本节分别从社会网络的整体性指标（表 5.8）、中心性与影响力指标、关联系与“小世界”指标进行了分析，以便发现网络中的特殊节点，以便分析个体网络演化与传播。

表 5.8 整体性网络指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
2.684	0.001	8	0.761	3.600	0.466

平均度：较小；图密度：该传播网络图比较稀疏；网络直径：较长，该网络的传播较远；模块化指标：划分出 7 个社区，且划分的质量较高；平均路径长度：长度较大，网络传播较广；平均聚类系数：较小。

## 2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算,这里取点度中心度排名前 6 的节点,列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名,得到对应的数据如表 5.9 所示。

表 5.9 中心度指标

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	值排名	绝对值	相对值	值排名	绝对值	相对值	值排名
Huazhong295	1008	0.282	1	198545	5.514	5	22522	8.427	179
fae79a84d7...	564	0.158	2	117119	3.253	8	20617	9.206	247
顿河骑兵	513	0.144	3	286411	7.955	1	16556	11.464	370
lixin330a	458	0.236	4	166827	4.633	6	19396	9.786	286
船老大的爱国心	421	0.118	5	204919	5.691	4	17541	10.820	298
8959685cc	73	0.068	6	0	0.000	-	3604302	0.053	701

从表 5.9 中可以看出“huazhong295”、“顿河骑兵”等 6 个节点的点度中心度排名前 6 位。从中间中心性来看,“顿河骑兵”的中间中心度远远大于其他节点,排在第一位,说明许多节点都需要经过该节点才能建立联系并获取信息。从接近中心性来看,“huazhong295”和“fae79a84d7...”的接近中心度排名前两位,说明这两个节点与其他节点进行交互比较容易。

图 5.13 的点度中心势指标为 0.28%,该指标值很小,表示该图在度数上向某点集中的趋势很不明显;该图的中间中心势指标为 7.92%,该指标值较大,表示该图向某些中间节点聚集的程度比较大,即在整个网络中大部分的结点需要经过某些中间节点才能得到信息。

## 3) 关联性与“小世界”

图 5.13 的特征途径长度为 5.949,凝聚度为 0.030。该图路径长度的具体分布如表 5.10 所示。

表 5.10 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	2696	0.005	5	140569	0.261
2	27094	0.050	6	95285	0.177
3	46437	0.086	7	45605	0.085
4	78838	0.146	8	23879	0.044

图 5.13 的聚类系数为 6.330,节点的聚类系数如表 5.11 所示。

表 5.11 聚类系数指标

用户	聚类系数	点对数
huazhong295	0.001	125250
fae79a84d7774c6e813c6ea5853515e3	0.003	44253
顿河骑兵	0.001	91806
lixin330a	0.001	27261
船老大的爱国心	0.003	31375
8959685cc06b43f4bbdf8c0a4a67a031	0.198	741

5.3.6 个体网络分析

在按照之前的整体网分析之后，我们从中选取那些节点度数相对较大、传播性比较好的帖子，做了某帖子下参与的所有用户的个体网络。本节选取的是“fae79a84d7774c6e813c6ea5853515e3”发表的一篇帖子，做可视图（图 5.14），对它进行指标分析，并描述传播路线。

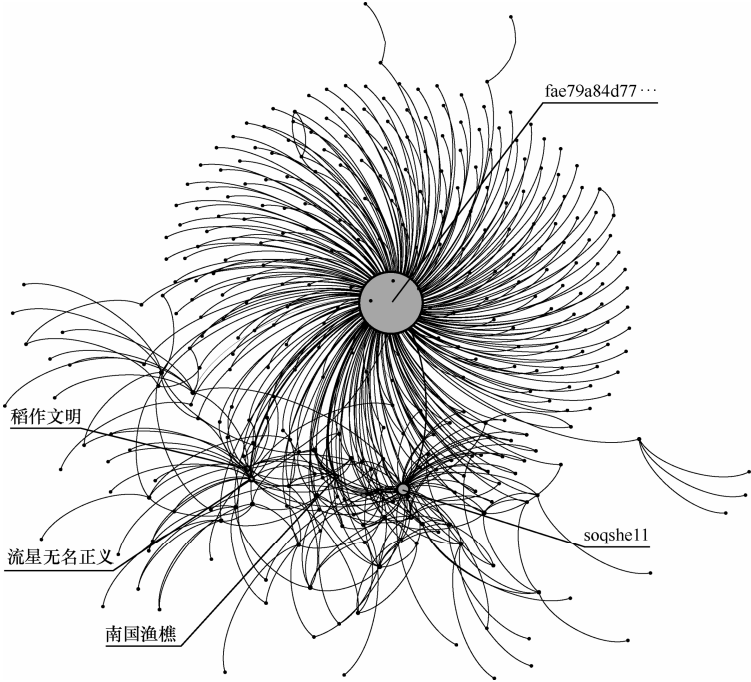


图 5.14 国企改革个体网络

1) 结构洞

这里列举节点度数排名前 5 位的用户的结构洞指标，形成相应的数据指标表（表 5.12）。



表 5.12 结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
fae79a84d77...	299	294.252	0.984	0.216	0.847
稻作文明	18	13.481	0.749	0.772	0.763
流星无名正义	54	43.115	0.798	0.310	0.646
子非鱼安知鱼乐	14	11.244	0.803	0.457	0.441
南国渔樵	11	4.346	0.395	0.843	0.617

从表 5.12 中可以看出“fae79a84d77...”的节点度数、有效规模和等级度都比较大,说明该居于网络的核心;同时该节点效率最大,限制度最小,表明该节点在网络中的行动最高效,受到网络中其他节点的限制比较小。“南国渔樵”对应的效率值较小,限制度较大,说明该节点在网络中受到其他节点的限制比较大,行动力较弱。

2) 诚实中间人

从表 5.13 中可以看出“流星无名正义”代表的节点作为中介的次数为 15 次,共联络了 105 对关系,其中 87 对关系属于 HBIO 类,14 对关系属于 HBIO1 类,4 对关系属于 HBIO2 类。这说明大部分节点属于纯中介,参与回复主贴的人物之间的关系比较弱。这也表明大部分回复是针对主贴回复的,而只有一小部分回复针对的是评论用户评论的。

表 5.13 诚实中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
fae79a84d7774c6e813c6ea5853515e3	4	6	5	0	1
soqshell	5	10	6	2	2
流星无名正义	15	105	87	14	4
子非鱼安知鱼乐	3	3	2	0	1

HBIO 代表纯中介, HBIO1 代表弱中介, HBIO2 代表非中介

从图 5.14 中可以看出“fae79a84d7774c6e813c6ea5853515e3”发表的帖子后大量用户参与的讨论。在回复的用户中,“流星无名正义”用户发表的评论引起了网民大量回复,说明该用户在网络传播中可能会成为一个意见领袖,他的意见会影响其他用户的思考和回复。

5.4 专题 2：反腐风暴

经过政治领域热点词处理以及政治领域意见领袖分析,选取反腐风暴作为

政治领域的第二个热点专题。“反腐”一词成为盘点 2014 年度的热词，突出表现了广大人民群众对于它的关注，习主席提出的“常、长”表示反腐要经常抓、反腐要长期抓，体现了党和政府对于反腐的态度。随着各方面体制机制的完善，相信反腐工作会让广大人民群众满意。

5.4.1 专题发展简介

腐败现象的滋生和蔓延，引起了广大人民群众的强烈不满。腐败的实质是公共权力的滥用，反腐败的核心是制约和权力监督。当前，滋生腐败的土壤依然存在，反腐斗争形势依然严峻复杂，一些不正之风和腐败问题影响恶劣，亟待解决。因此，应坚持和强化法治思维、法治方法，加快推进依法治国，不断完善惩治和预防腐败的法律法规制度体系，用法制约权力，在法治框架下反对腐败，使反腐败走向规范化、制度化。特别是要加强改革权限，用法治压缩腐败滋生空间。

腐败问题不仅成为民众最为痛恨的问题，它也是我党必须解决的关键性问题。腐败问题高发的根源在于权力过大而得不到有效制约，如何推进反腐倡廉工作，重点还在于坚持标本兼治、综合治理。惩防并举、注重预防的方针，建立惩治和预防腐败体系，加强权力运行制约，让人民来监督权力。因此，解决腐败问题的根本出路在于公共权力的大幅退缩。特别是在政治体制改革的背景下，对政府与公权力进行合理与有效的监督成为关键。

坚持“老虎苍蝇一起打”的中央纪律检查委员会（中纪委）在 2014 年诠释了一副壮观的反腐图，637 名落马“贪官”被指名道姓通报。互联网上，广发网民对于“反腐风暴”展开了激烈的讨论（表 5.14）。

表 5.14 反腐风暴专题热帖

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
渐清的吏治，依稀见到雍正时代的影子！	cat_alx	2014/04/12	290992	3043	725	894	52.540
一个副科级干部的敛财之道(转载)	赶大的人	2014/03/26	8335	484	158	197	32.547
【普欣夜话】官员泛腐极贫之根，品德乎，制度乎？	d07150...	2014/06/21	53480	422	188	111	3.540
小小村委主任竟贪过亿，花都狮岭真是沃土(续)!!!	普通一村民	2014/01/05	23969	213	84	94	27.654
三观尽毁！奥巴马表示：美国民主是合法的腐败！	每天一顿饭	2014/11/21	38607	324	126	95	11.833

### 5.4.2 观点分类与观点列表

以论坛中网民的回复为研究对象,通过观点挖掘和语义分析形成本专题内的核心观点,并对其进行分类和详细描述。

#### 1) 观点分类

根据观点涉及的特征词汇和领域以及网民发表的各个观点内容本身之间存在的相关性与差异性,可以将所有评论分为以下三个主题。

##### 主题一:当前腐败发生“势不可当”,原因也是多方面的

此分类观点包含四个关键特征词:腐败、监督、制度、法律。对应的观点包括:

观点1. 对权力的监督机制不完善、监督力度不大。

观点2. 国家基层人员工资偏低导致腐败。

观点3. 人们的法律意识淡薄,道德沦丧。

##### 主题二:防止腐败的发生要从多方面入手

此分类观点包含四个关键特征词:腐败、惩治、法治、道德。对应的观点包括:

观点4. 加强思想道德建设和监管机制来预反腐败。

观点5. 腐败官员的惩治措施要严肃、公开、透明。

观点6. 提高公职人员待遇,高薪养廉。

观点7. 反腐与扫黄、抓赌、清毒等同步进行,它们有千丝万缕的联系。

##### 主题三:反腐政治的“衍生物”,与百姓无关,只是看个热闹

此分类观点包含三个关键特征词:反腐、百姓、政治。对应的观点包括:

观点8. 只要有官员因为反腐落马,拍手称快,有“仇官”的即视感。

观点9. 反腐与政治的关系千丝万缕。

从图 5.15 中可以看出,36%的网民提出的观点绝大多数集中在当前腐败现象的原因讨论上;37%的网民期望通过建立一个法治社会来降低和抑制腐败的发生;27%的网民对于反腐问题不太关心,或认为与自己没关系,或认为反腐是“政治游戏”。这说明大部分网民对于反腐的“参与感”较强,“参与度”较高,只有少部分网民对于政府反腐“漠不关心”。这是一种积极的信号,希望党和政府真的可以在反腐工作中取得显著成绩,“对得起”广大网民的关注与参与。

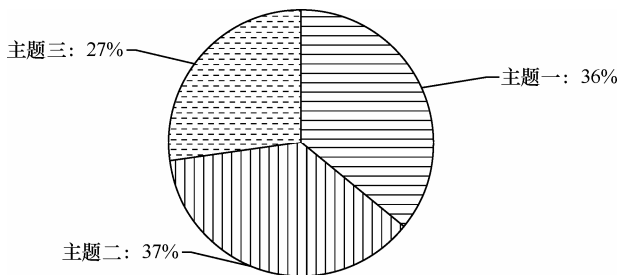


图 5.15 “反腐风暴”专题观点分类比例图

## 2) 观点分类详细分析

为了更好地从宏观向微观的深入分析，本节对每个主题内的网络用户分布以及网络民意进行深入分析，希望通过网络用户真实的反馈信息来反映他们对这些问题的思考与建议。

### 主题一：当前腐败发生“势不可当”，原因是多方面的

在进一步的数据分析与研究中，对网民在这个主题内的深入梳理与分析，发现存在 2 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 51% 的评论认为对权力的监督机制不完善、监督力度不大是发生腐败的主要原因。

例如，“加大监管，支持反腐”“力挺国家的反腐，监管”。

实践表明，失去监督的权利，必然导致腐败现象的发生。没有有效的监督，就不会有对腐败现象的真正遏制。权力腐败的实质是滥用权力。相信监管机制的完善会对反腐工作提供有力的保障。

观点 2. 8% 的评论认为国家基层人员工资偏低导致腐败。

例如，“高薪养廉，这是老祖宗们用过的法子”。

公务员的低工资似乎缩短了政府工作人员与人民之间的距离，实现了反对特权和实现公正的目的。但是，实际上在亚非拉等发展中国家，政府官员的滥用权力、贪污受贿等行为正好成为低工资的一种“合理”而不合法的反例。

观点 3. 41% 的评论认为人们的法律意识淡薄、道德沦丧。

例如，“全面积腐败，没有什么领域是干净的。改革开放三十年需要反思的东西很多，任何一个时代，有成绩，有脓包，这全很正常。反思改革开放三十年很重要。改革开放三十年来，无冕之王记者变成了鸡，白衣天使的医生变成了吃人的狼……社会道德沦丧！”

当今社会，由于道德理想的迷失、道德观念的错位、道德准则的失范，引起了道德行为的失序。这种失序反应在政治生活方面便是权力的腐败、贪污受

贿现象的发生。更为严重的是,许多人认为贪污是公务员的基本属性,并且默认了这种属性,这是极端错误的。

综上,在本主题中,具有相应子观点的评论分布比例如图 5.16 所示。

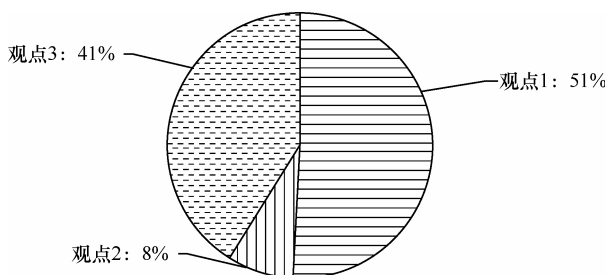


图 5.16 主题一细分子观点比例图

## 主题二：防止腐败的发生要从多方面入手

在进一步的数据分析与研究中,对网民在这个主题内的深入梳理与分析,发现存在 4 个细分的子观点。这些细分的子观点所反映的内容如下:

观点 1. 46%的评论认为加强思想道德建设和监管机制来预防腐败。

例如,“加大监管力度,坚决抵制不作为的官员”。

加强政治伦理建设时,要弘扬优良的德政传统,唤醒权力主体的自律意识,培养廉洁自律的政治氛围。加强思想道德建设的同时,还要加强监督机制建设,监督保廉,从而建立一个与市场经济发展相适应的权力运行机制。

观点 2. 28%的评论表示腐败官员的惩治措施要严肃、公开、透明。

例如,“零容忍态度惩治腐败!顶起来!”

对于贪腐官员的处理问题,目前还存在诸多问题,甚至有“顶包”进监狱一说。对于腐败官员处置的相应要求,其实也是对法治社会的基本要求。

观点 3. 6%的评论表示应该提高公职人员待遇,高薪养廉。

例如,“基层普通公务员确实收入微薄,大部分不说兢兢业业,也是履行了职责的,对他们来说收入和付出不相称,有意见很正常,国家应当考虑正常的利益诉求。”

观点 4. 20%的评论认为反腐与扫黄、抓赌、清毒等同步进行,它们有千丝万缕的联系。

例如,“现在的政府任重道远,边完善制度边打老虎拍苍蝇。反反腐,扫扫黄,抓抓赌,清清毒,一个都不落。”

那些贪官们“最喜欢”去的赌场、色情服务场所被清理干净后,对于贪污

现象肯定会有所帮助。但是,根本上还是要依靠思想道德教育和法治监管。

综上,在本观点中,具有相应子观点的评论分布比例如图 5.17 所示。

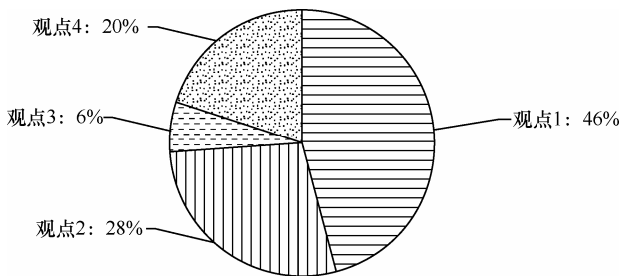


图 5.17 主题二细分子观点比例图

### 主题三：反腐政治的“衍生物”，与百姓无关，只是看个热闹

在进一步的数据分析与研究中,对网民在这个主题内的深入梳理与分析,发现存在 2 个细分的子观点。这些细分的子观点所反映的内容如下:

观点 1. 46%的评论认为只要有官员因为反腐落马,拍手称快,有“仇官”的即视感。

例如,“中国的官员,只要涉及权力范围,无官不贪无官不腐,则是一定的!”

许多网友对于因为反腐落马的官员第一反应就拍手称快,似乎所有官员都在贪污,所有官员都应该落马,仅仅是落马时间的问题。但是,“仇官”的想法并不能帮助社会解决反腐的问题。

观点 2. 54%的评论认为反腐与政治的关系千丝万缕。

例如,“就是政治洗牌,排除异己,反腐就是一顶帽子和当年反革命一样。等换上去的都是自己人就消停了。”

诸如此类的网友版反腐解读似乎很有道理,但是经不起仔细推敲。落马官员无不是贪污腐败所造成,与网友们所说的政治似乎关联不大,还需正确对待。

综上,在本主题中,具有相应子观点的评论分布比例如图 5.18 所示。

### 3) 观点总结

腐败现象的发生实质是公共权力的滥用,反腐败的核心是制约和监督权力。当前社会中滋生腐败的土壤依然存在,反腐败斗争形势依然严峻,一些不正之风和腐败问题影响恶劣,亟待解决。

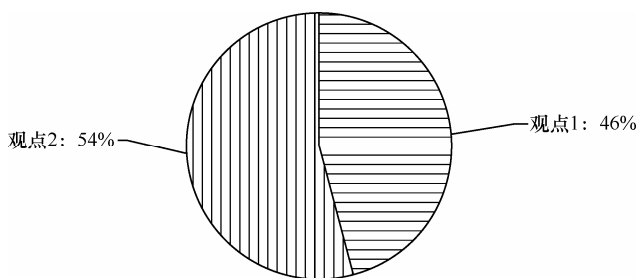


图 5.18 主题三细分子观点比例图

因此，一方面应坚持和强化法治思维、法治方式，加快推进依法治国，不断完善惩治和预防腐败的法律法规制度体系，用法制约束权力，在法治框架下反对腐败，使反腐败走向规范化、制度化。另一方面，腐败问题高发的根源在于权力过大而得不到有效制约，解决腐败问题的根本出路在于公共权力的大幅退缩以及监督制约机制的建立。政府必须以法制规范权力，以民主监督权力，建立并完善以法律控制权力、以权力制约权力的制度和机制。通过规范运行流程，使隐性权力公开化，显性权力规范化。在法律法规的约束下，做到决策权、执行权、监督权既相互制约又相互协调，实现科学分权与相互制衡。

此外，对于广大群众的思想认识和反腐教育还有较长的路要走，要将广大群众发动起来，利用网络科技和信箱等手段，让人民行使监督权力，让权力在阳光下运行，最大限度地减少腐败问题滋生的体制障碍和制度漏洞。

### 5.4.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

#### 1) 影响力

专题的传播过程中流量是指相关帖子的回复数量，发展峰值出现在舆论讨论最激烈的时间点。很明显，流量与影响力成正比关系，即流量越大影响力也就越大。因此，本节分别对新增回复数和总回复数按照时间进行统计，然后对统计结果进行可视化得到波线图（图 5.19，图 5.20），以便发现专题传播的发展峰值。

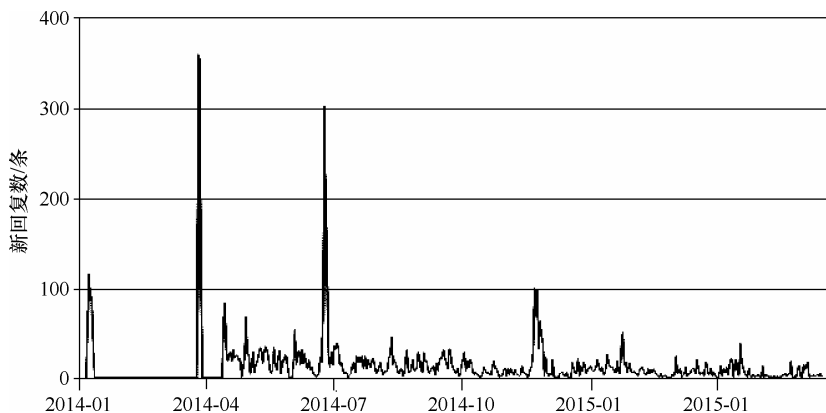


图 5.19 专题日流量时间趋势图

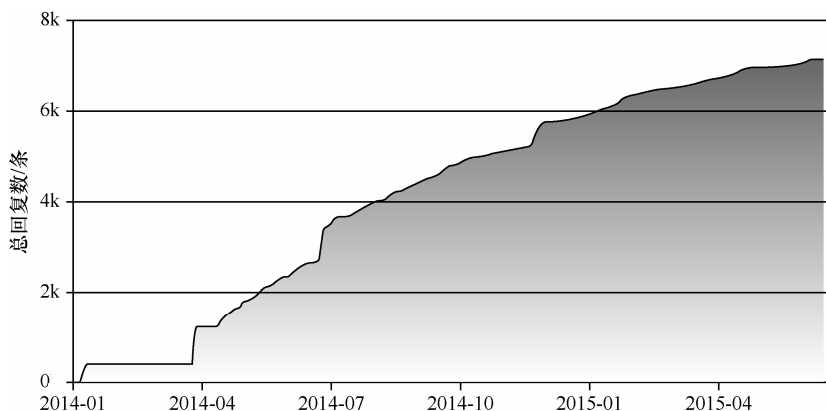


图 5.20 专题总流量时间趋势图

从图 5.19、图 5.20 中可以看出，此专题的流量集中在 2014 年 1 月 6 日、2014 年 3 月 27 日、2014 年 11 月 22 日。在专题日流量中这几天的新回复数量较大，在专题总流量中这几天回复数的增速较大。反腐风暴专题在 3 月 27 日的日流量达到最大值，回复值为 360；3 月 27 日过后，该专题的总流量持续保持增长。这说明 3 月 27 日广大网友交流较多，该天的话题同时引起了后续的讨论，在此次讨论没有结束时，该专题下的新帖又引起了网友的热议，导致该专题的总流量持续增加。

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库，对专题内的所有回复内容进行情感分析。情感分析包括对专题内情感总体偏向（图 5.21）的分析和专题内回复内容的情感分类（图 5.22）的分析。



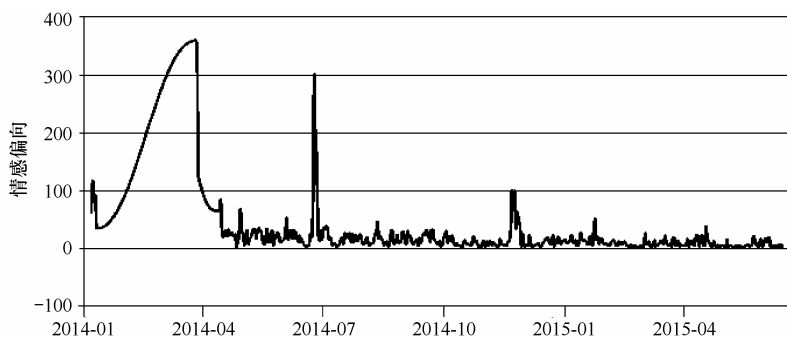


图 5.21 专题情感趋势图

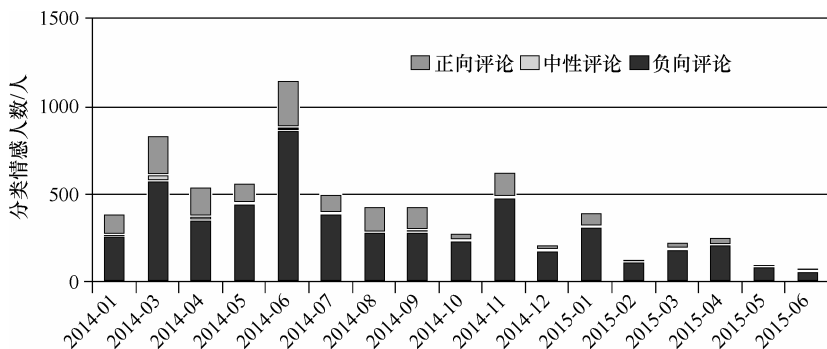


图 5.22 专题情感分类图

从图 5.21 可以看出，发展峰值出现在 2014 年 3 月 26 日、2014 年 6 月 24 日、2014 年 5 月 21 日、2014 年 11 月 28 日，事件情感值较大，同样这几天所对应的日流量数值较大。因为较大的日流量才会使当日的情感偏向波动较为明显，较小的日流量会使情感偏向趋向于零。然而，并不是说有较大的日流量一定会出现情感偏向的大幅波动，较大的日流量情感偏向也有可能是正常的。

从图 5.22 中可以发现，负向评论的数量大于正向评论和中性评论，这是由于网络环境下语言的随意性和较低的约束性造成的。甚至有些月份几乎没有正向评论，可以理解这种现象的发生，毕竟许多网友上网是为了释放情绪与缓解压力。

#### 5.4.4 专题用户情感的分析

随着近十年来论坛的兴起，网民数量也急剧上升，网民作为网络舆情的一个非常重要因素，行为和言论会引导整个舆论的走向。而在网民中那些发言多、回复多的“意见领袖”或者“领军人物”，往往会主导整个舆论的走向。所以说，探索专题下的人物的情感可以帮助我们发现该专题下的情感正向用户，同时也可以帮助论坛找出到处散发消极影响的网民，剔除该用户以净化论坛环境。

本节对各个专题下的人物情感进行了统计分析, 由于活跃度对情感的影响较大, 所以本节构建了分别以情感值和活跃度为两个坐标轴的专题下任务的散点图 (图 5.23)。

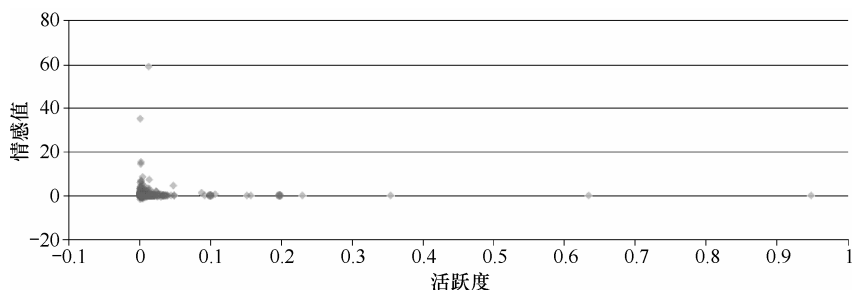


图 5.23 基于活跃度和情感值的用户散点图

圆点代表发帖人, 方点代表评论人

从图 5.23 中可以看出, 一般发帖人的情感值较高, 那些活跃度较高同时情感值较高的用户可以被视为“意见领袖”。同时存在一些不是发帖者但其活跃度较高的, 可能该用户所评论的内容引发了其他网民的热烈讨论, 也可能就是该用户经常在论坛中灌水、评论等, 具有较高的回复参与量, 但是通常回复的内容情感极性较低, 感情色彩不明显, 但是这些用户所属领域性比较明显。

为了进一步进行分析, 本节将那些情感值负性较大且活跃度较大的用户进行了详细点的描绘, 取情感值小于 20、活跃度大于 0.02 的所有用户, 作图 5.24。这是一类在论坛中活动积极、负面情绪较大的用户。这类用户的帖子和评论影响到其他人对国企改革看法。

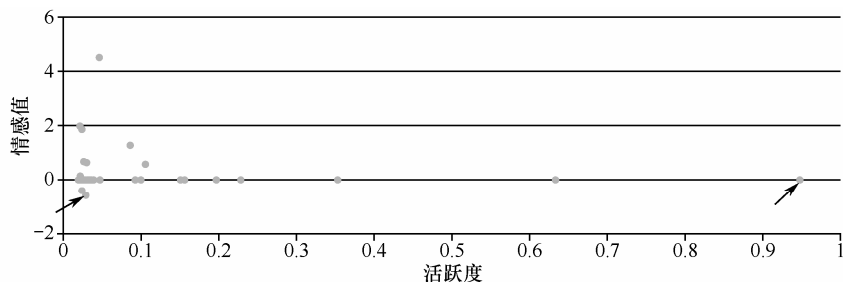


图 5.24 低情感和高活跃度散点图

从图 5.24 可以观察到, 那些活跃度较高同时情感值较低的用户, 论坛管理员要对这些用户进行监察和历史查询, 当这些用户出现不正当言论是进行及时处理。

5.4.5 专题网络的构建与分析

构建专题网络的方法见节 4.3.5。构建出股市与股票的专题网络，进行可视化，结果见图 5.25。

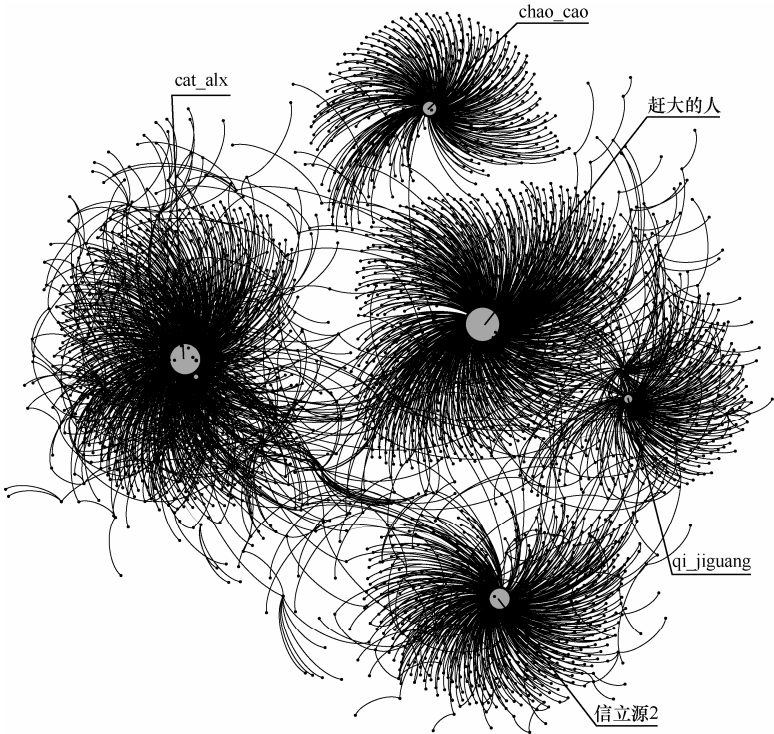


图 5.25 专题传播网络

由图 5.25 可知，图中有五个较大的社区，选择“cat\_alx”和“信立源 2”发布的帖子及受众构成的社区进行分析。

1) 网络指标概述

本节分别从社会网络的整体性指标（表 5.15）、中心性与影响力指标、关联系与“小世界”指标进行了分析，以便发现网络中的特殊节点，以便分析个体网络演化与传播。

表 5.15 整体性网络指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
2.508	0.001	10	0.598	4.464	0.584

平均度：较小；图密度：该传播网络图比较稀疏；网络直径：较长，该网络的传播较远；模块化指标：划分出 5 个社区，且划分的质量较高；平均路径长度：长度较大，网络传播较广；平均聚类系数：较小。

## 2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算, 这里取点度中心度排名前 5 的节点, 列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名, 得到对应的数据如表 5.16 所示。

表 5.16 中心度分析

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	值排名	绝对值	相对值	值排名	绝对值	相对值	值排名
信立源 2	1849	0.126	1	99420	1.248	7	4789542	0.059	13
cat_alx	1504	0.102	2	354037	4.444	1	4788993	0.059	11
qi_jiguang	1023	0.070	3	136021	1.707	3	4789792	0.059	14
chao_cao	942	0.064	4	166931	2.095	2	4789803	0.059	15
赶大的人	789	0.054	5	0	0.000	-	5637581	0.050	410

注：一代表排名非常靠后。

从表 5.16 中可以看出“信立源 2”“cat\_alx”等 5 个节点的点度中心度排名前 6 位。从中间中心性来看, “cat\_alx”的中间中心度远远大于其他节点, 排在第一位, 说明许多节点都需要经过该节点才能建立联系并获取信息。从接近中心性来看, “cat\_alx”和“信立源 2”的接近中心度排名前两位, 说明这两个节点与其他节点进行交互比较容易。

图 5.25 的点度中心势指标为 0.13%, 该指标值很小, 表示该图在度数上向某点集中的趋势很不明显; 该图的中间中心势指标为 4.44%, 该指标值较小, 表示该图向某些中间节点聚集的程度比较小, 即在整个网络中大部分的结点并不需要经过某些中间节点, 就能得到信息。

## 3) 关联性与“小世界”

图 5.25 的特征途径长度为 3.802, 凝聚度为 0.018。该图路径长度的具体分布如表 5.17 所示。

表 5.17 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	3704	0.008	5	47393	0.100
2	68354	0.145	6	70985	0.150
3	197461	0.418	7	17024	0.036
4	64585	0.137	8	2554	0.005

图 5.25 的聚类系数为 18.306, 节点的聚类系数如表 5.18 所示。

表 5.18 聚类系数指标

用户	聚类系数	点对数
信立源 2	0.475	946
新乡阿剑	0.000	20100
神奇的超短线	0.003	29646
牛牛依旧	0.000	60378
火哥闯天下	0.000	101025

5.4.6 个体网络分析

在按照之前的整体网分析之后，我们从中选取那些节点度数相对较大、传播性比较好的帖子，做了某帖子下参与的所有用户的个体网络。本节选取的是“cat\_alx”发表的一篇帖子，做可视图（图 5.26），对它进行指标分析，并描述传播路线。

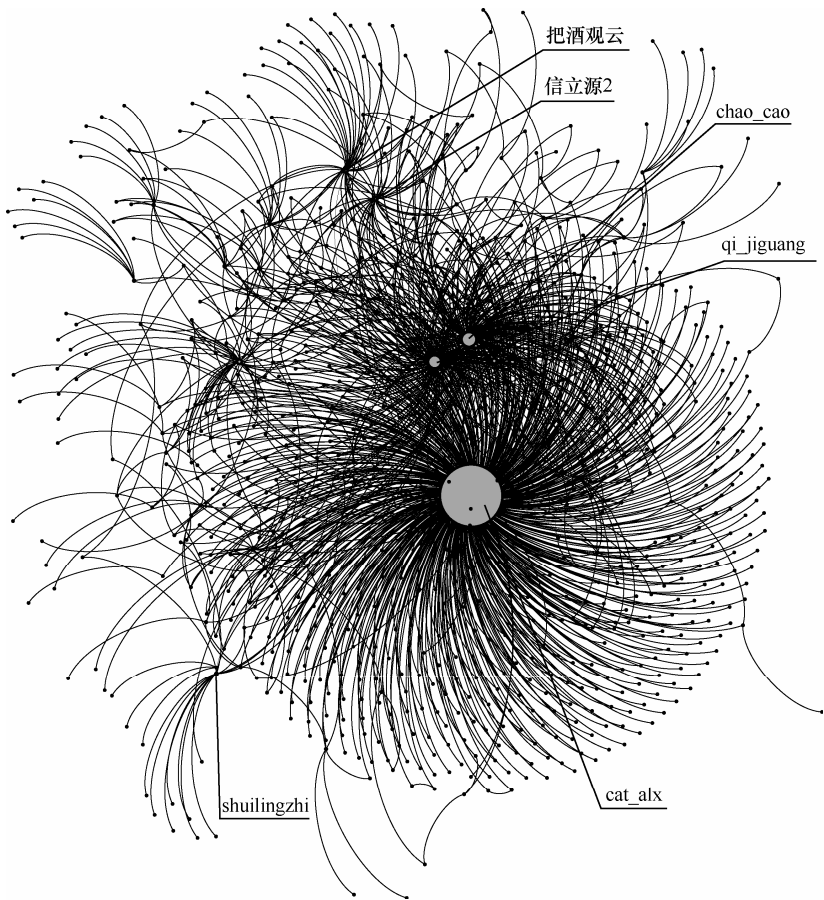


图 5.26 反腐风暴个体网络

从图 5.26 中可以看出,“cat\_alx”发表的帖子后,大量用户参与的讨论,在回复的用户中,“chao\_cao”和“qi\_jiguang”发表的评论引起了网民大量回复。这说明该用户在网络传播中可能会成为一个意见领袖,他的意见会影响其他用户的思考和回复。

1) 结构洞

这里列出节点度数排名前 5 位的用户的结构洞指标,形成相应的数据指标表(表 5.19)。

表 5.19 结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
信立源 2	45	31.004	0.689	0.481	0.590
chao_cao	159	135.081	0.850	0.626	0.780
cat_alx	710	663.949	0.935	0.241	0.809
把酒观云	3	2.167	0.722	1.409	0.293
shuilingzhi	14	13.750	0.982	0.100	0.064

从表 5.19 可以看出“cat\_alx”的节点度数、有效规模和等级度都比较大,说明该居于网络的核心;“shuilingzhi”效率最大,限制度最小,表明该节点在网络中的行动最高效,受到网络中其他节点的限制比较小。“把酒观云”对应的效率值较小,限制度较大,说明该节点在网络中受到其他节点的限制比较大,行动力较弱。

2) 诚实中间人

从表 5.20 中可以看出,“cat\_alx”代表的节点作为中介的次数为 72 次,共联络了 2556 对关系,其中 194 对关系属于 HBIO 类,194 对关系属于 HBIO1 类,17 对关系属于 HBIO2 类。这说明大部分节点属于纯中介,参与回复主贴的人物之间的关系比较弱。这也表明大部分回复是针对主贴回复的,而只有一小部分回复针对的是评论用户评论的。

表 5.20 诚实中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
信立源 2	13	78	53	15	10
chao_cao	17	136	109	16	11
cat_alx	72	2556	194	194	17
qi_jiguang	19	171	140	16	15

HBIO 代表纯中介,HBIO1 代表弱中介,HBIO2 代表非中介。

5.5 专题 3：依法治国

经过政治领域热点专题热点词处理以及政治领域意见领袖分析，选取依法治国作为政治领域的第三个热点专题。依法治国就是依照体现人民意志和社会发展规律的法律治理国家，而不是依照个人意志、主张治理国家；要求国家的政治、经济运作、社会各方面的活动统统依照法律进行，而不受任何个人意志的干预、阻碍或破坏。依法治国是中国共产党领导人民治理国家的基本方略，是发展社会主义市场经济的客观需要，是社会文明进步的显著标志，是国家长治久安的必要保障，是建设社会主义法治国家、人民当家做主根本保证。

5.5.1 专题发展简介

经过改革开放三十多年的法治发展，中国特色社会主义法律体系已经形成。法律已成为约束人们行为的规范和维护社会秩序的必要手段，目前我国在社会活动各方面基本实现“有法可依”的条件下，法治国家建设的主要任务将从加强立法转向更加注重法的实施，强调通过严格执法公正司法和全民守法实现建设法治中国的总体目标。

依法治国是一项立法、执法、司法和守法共同推进的综合性工程，需要法治国家、法治政府和法治社会的一体化建设。中国共产党第十八届中央委员会第四次全体会议（中国共产党十八届四中全会）通过的《中共中央关于全面推进依法治国若干重大问题的决定》提出了建设中国特色社会主义法治体系，希望每个案件均可以实现社会的公平正义，积极回应社会各界对司法公正和依法治国的关注和期待。在互联网上，广大群众对于“依法治国”展开了激烈的讨论（表 5.21）。

表 5.21 依法治国专题热帖

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
四中全会如沐春风，依法治国	易天校园 2011	2014/10/25	53298	805	500	168	99.254
一个副科级干部的敛财之道（转载）	赶大的人	2014/03/26	8s335	484	158	197	32.547
【老墨杂弹】黄秋生被封杀的中国法治 环境	d61a6ae03...	2014/11/08	140161	1072	352	242	98.619
河南联通平顶山分公司：员工请假被除名视劳动法如玩具	OK 千里千寻	2014/09/21	825089	1046	249	439	0.400
党的群众路线教育实践活动，能保群众不受政府职能部门伤害不？	新手 L	2014/02/19	626817	1948	382	812	1.390

## 5.5.2 观点分类与观点列表

本节以论坛中网民的回复为研究对象，通过观点挖掘和语义分析形成本专题内的核心观点，并对其进行分类和详细描述。

### 1) 观点分类

根据观点涉及的特征词汇和领域以及网民发表的各个观点内容本身之间存在的相关性与差异性，可以将所有评论分为以下三个主题。

#### 主题一：对目前依法治国现状的描述

此分类观点包含三个关键特征词：执法、监督、法律。对应的观点包括：

观点1. 我国目前处于法治社会初级阶段，执法不严，徇私枉法。

观点2. 违法的代价太小，有法不依现象严重。

观点3. 法律不完备，存在真空。

#### 主题二：对我国目前法治社会发展的评价

此分类观点包含三个关键特征词：进程、法治、体制。对应的观点包括：

观点4. 依法治国建设进程过慢，跟不上社会的发展速度。

观点5. 法治社会的建设是劳动人民的共同理想。

观点6. 我国的司法体制正在进步、走向好转，法律在逐渐被人们认同。

#### 主题三：依法治国的建议和意见

此分类观点包含2个关键特征词：群众、社会监督。对应的观点包括：

观点7. 法律法规的制定需要精英阶层，但是普通民众也需要参与其中。

观点8. 加强社会监督，避免钱权交易建立在法律之上。

从图5.27可以看出，许多网民表示目前我国司法体制中存在诸多问题，甚至有部分网民对目前的司法体制产生不信任的感觉。36%网民是在描述目前依法治国的现状及评价，对于依法治国的建议者数量较少。这说明广大网民都积极参与者法治社会的建设，但是由于自身学识和知识储备的原因，没有能力提出关于依法治国的相关建议。

依法治国首先要有法可依，目前我国的司法制度不完善，网民关心司法改革能不能改善人们的切身利益。希望新阶段的政治体制改革强化司法改革的力量与相关的顶层设计，着重实现社会的公平和正义。

### 2) 观点分类详细分析

为了更好地从宏观向微观的深入分析，本节对每个分类观点内的网络用户分布以及网络民意进行了深入分析，希望通过网络用户真实的反馈信息来反映他们对这些问题的思考与建议。



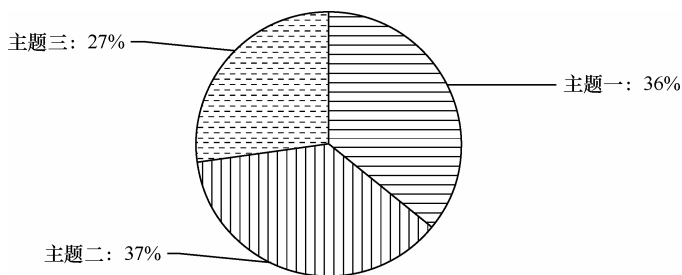


图 5.27 “依法治国”专题观点分类比例图

### 主题一：对目前依法治国现状的描述

在进一步的数据分析与研究中，对网民在这个主题内的深入梳理与分析，发现存在 3 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 45% 的评论认为我国目前处于法治社会初级阶段，执法不严，徇私枉法时有发生。

例如，“依法治国，国民之幸。希望能从上到下都是这样！不要让有些人照样目无法纪，我行我素！”

我国的法制现代化进程从头算起不过百年，改革开放以来的社会主义法制现代化建设也只有三十多年，初级阶段是存在的，但肯定是可以过去的。

观点 2. 36% 的评论认为违法的代价太小，有法不依现象严重。

例如，“弱势群体遭遇有法不依的悲凉处境何时能结束？”“法治社会，有这样的有法不依的法官，我们很失望啊”。

公务员素质的参差不齐是制约依法治国的重要因素，提高公务员的法律常识与个人素质迫在眉睫。

观点 3. 19% 的评论认为法律不完备，存在真空。

例如，“你错了，这个问题错在法律不完善，执法不严谨，证据不充分。如果法律上对老人碰瓷敲诈这个问题进行惩罚，你看还有老人敢碰瓷嘛！如果有，碰瓷后他就要坐牢，赔钱！让好心人的付出不被抹黑，还有人不去扶起老人嘛？那个法官你怎么可以说他没有道德呢？你无法判断。也许他的道德标准和你不一样呢？”

随着社会的发展，许多新鲜事物横空出世，旧的法律不可能包括全面，必然存在真空，依法治国的推进还是需要先进行法治的改革。

综上，在本主题中，具有相应子观点的评论分布比例如图 5.28 所示。

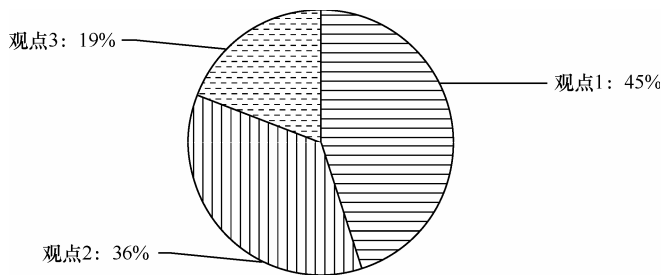


图 5.28 主题一细分子观点比例图

### 主题二：对我国目前法治社会发展的评价

在进一步的数据分析与研究中，对网民在这个观点内的深入观点进行梳理与分析，发现存在 3 个细分子观点。这些细分子观点所反映的内容如下：

观点 1. 37% 的评论认为依法治国建设进程过慢，跟不上社会的发展速度。

例如，“依法治国，首先把法律完善……现在法律漏洞百出，有的十几年都没有修改……法律成摆设。当官的一批示、一关注，比什么法律都强……”

社会发展过快，法律的修订不及时，导致依法治国的建设跟不上社会的发展速度。

观点 2. 35% 的评论表示法治社会的建设是劳动人民的共同理想。

例如，“依法治国，依法维护人民权益、维护社会公平正义、维护国家安全稳定，为实现中华民族伟大复兴的中国梦提供有力法治保障。”

随着法律制度的不断完善，人民教育水平和法律知识的不断提高，大部分人还是希望推进依法治国，生活在一个法治社会中。

观点 3. 28% 的评论表示我国的司法体制正在进步、走向好转，法律在逐渐被人们认同。

例如，“基层普通公务员确实收入微薄，大部分不说兢兢业业，也是履行了职责的，对他们来说收入和付出不相称，有意见正常，国家应当考虑正常的利益诉求。”

中国向法制建设迈出重要一步到现在短短三十多年的时间，想要西方国家上百年的立法进程取得的成果很不现实，立足当下，让更多的人认同法律，早日实现依法治国。

综上，在本观点中，具有相应子观点的评论分布比例如图 5.29 所示。

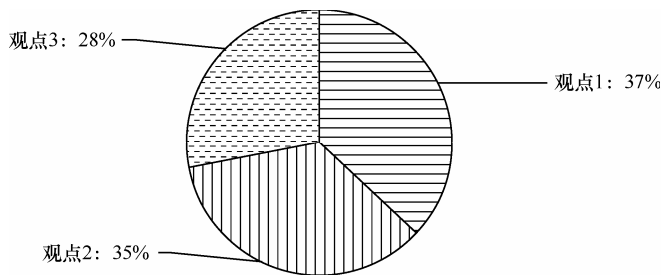


图 5.29 主题二细分子观点比例图

### 主题三：依法治国的建议和意见

在进一步的数据分析与研究中，对网民在这个主题内的深入梳理与分析，发现存在 2 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 61%的评论认为法律法规的制定需要精英阶层，但是普通民众也需要参与其中。

例如，“对制度的尊重，对法律的信赖，才是国家和民众的未来”。

实现依法治国，不仅仅需要精英阶层的参与，普通民众的法律意识都需要提高。只有全民懂法，才会实现依法治国。

观点 2. 39%的评论认为加强社会监督，避免钱权交易建立在法律之上。

例如，“如果权力法律没有强势的舆论监督，权力法律就会成为某些人的私人‘机器’。”

早日实现依法治国，必须加强社会监督，采用有建设性的建议，避免钱权交易建立在法律之上。政府应该鼓励新闻媒体对各种违法乱纪现象的报道和曝光，确保新闻自由对滥用权力的制约。

综上，在本主题中，具有相应子观点的评论分布比例如图 5.30 所示。

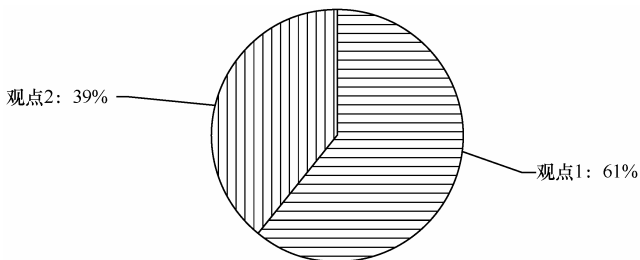


图 5.30 主题三细分子观点比例图

### 3) 观点总结

目前，依法治国的实施存在两个较大问题。首先是我国司法体系中存在诸多不尽如人意的地方，许多问题没有得到很好地解决。特别是当前我国处在社

会转型期，社会矛盾尤为突出，整治好这些问题就必须做好法制现代化的工作，真正发挥好司法机关的作用。另一个问题，就是目前许多国民不具备基本法律素质，受教育程度较低，如何将法律意识根植于人心、如何提高民众法律意识是亟待解决的问题，同时也是实现依法治国的基础。再健全的司法体制，少了民众的支持也不会发挥其应有的作用。

### 5.5.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

#### 1) 影响力

专题的传播过程中流量是指相关帖子的回复数量，发展峰值出现在舆论讨论最激烈的时间点。很明显，流量与影响力成正比关系，即流量越大，影响力也就越大。因此，本节分别对新增回复数和总回复数按照时间进行统计，然后对统计结果进行可视化得到波线图（图 5.31，图 5.32），以便发现专题传播的发展峰值。

从图 5.31、图 5.32 中可以看出，此专题的流量集中在 2014 年 2 月 22 日、2014 年 7 月 27 日、2014 年 11 月 10 日。在专题日流量中这几天的新回复数量较大，在专题总流量中这几天回复数的增速较大。依法治国专题在 2 月 22 日的日流量达到最大值，回复值为 820。7 月 27 日的话题引起了后续的讨论，在此次讨论没有结束时，该专题下的新帖又引起了网友的热议，导致该专题的总流量在 7 月过后持续增加至 12 月。

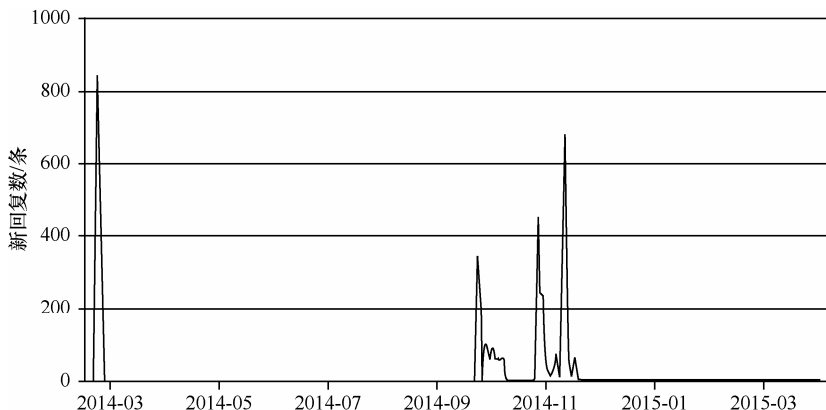


图 5.31 专题日流量时间趋势图

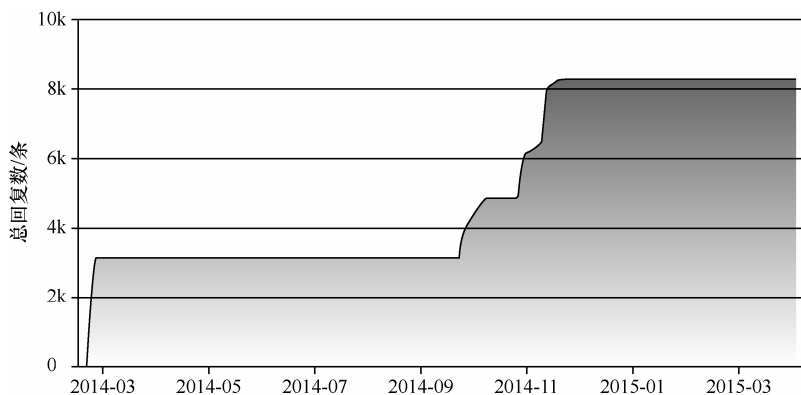


图 5.32 专题总流量时间趋势图

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库，对专题内的所有回复内容进行情感分析。情感分析包括对专题内情感总体偏向（图 5.33）的分析和专题内回复内容的情感分类（图 5.34）的分析。

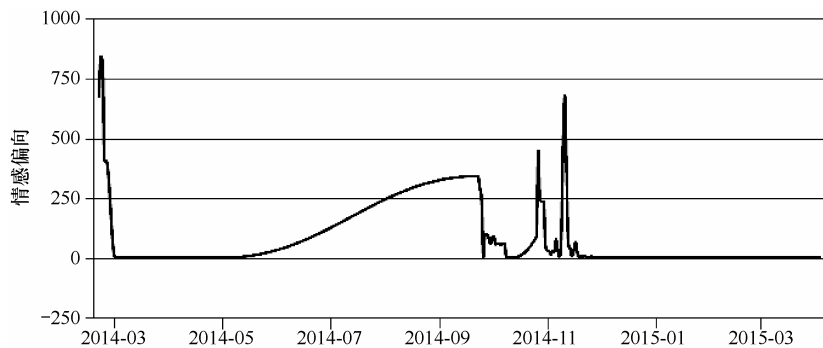


图 5.33 专题情感趋势图

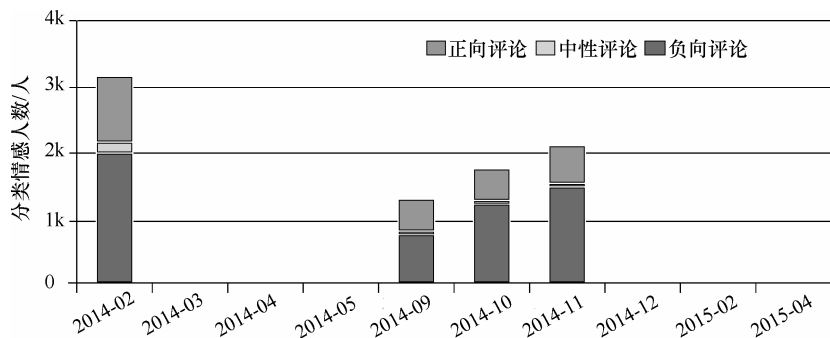


图 5.34 专题情感分类图

从图 5.33 可以看出，发展峰值出现在 2014 年 2 月 22 日、2014 年 7 月 23 日、2014 年 10 月 26 日、2014 年 11 月 10 日，事件情感值较大，同样这几天所对应的日流量数值较大。因为较大的日流量才会使当日的情感偏向波动较为明显，较小的日流量会使情感偏向趋向于零。然而并不是说有较大的日流量一定会出现情感偏向的大幅波动，较大的日流量情感偏向也有可能是正常的。

从图 5.34 中可以发现，负向评论的数量远大于正向评论和中性评论，这是由于网络环境下语言的随意性和较低的约束性造成的。

### 5.5.4 专题用户情感的分析

随着近十年来论坛的兴起，网民数量也急剧上升，网民作为网络舆情的一个非常重要因素，行为和言论会引导整个舆论的走向。而在网民中那些发言多、回复多的“意见领袖”或者“领军人物”，往往会主导整个舆论的走向。所以说，探索专题下的人物情感可以帮助我们发现该专题下的情感正向用户，同时也可以帮助论坛找出到处散发消极影响的网民，剔除该用户以净化论坛环境。

本节对各个专题下的人物情感进行了统计分析，由于活跃度对情感的影响较大，所以本节构建了分别以情感值和活跃度为两个坐标轴的专题下任务的散点图（图 5.35）。

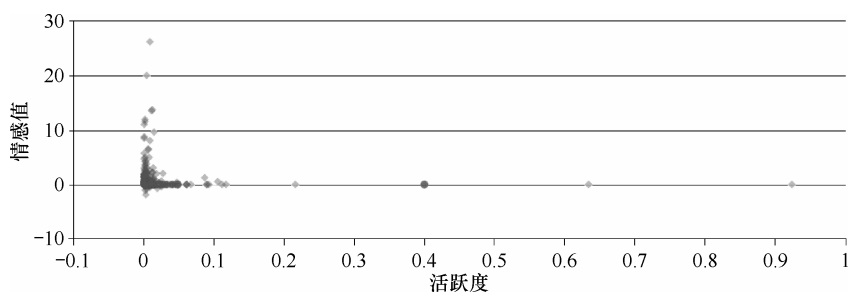


图 5.35 基于活跃度和情感值的用户散点图

圆点代表发帖人，方点代表评论人

从图 5.35 中可以看出，一般发帖人的情感值较高，那些活跃度较高同时情感值较高的用户可以被视作意见领袖。同时存在一些不是发帖者但其活跃度较高的点，可能该用户所评论的内容引发了其他网民的热烈讨论，也可能

就是该用户经常在论坛中灌水、评论等，具有较高的回复参与量，但是通常回复的内容情感极性较低低，感情色彩不明显，但是这些用户所属领域性比较明显。

为了进一步进行分析，本节将那些情感值负性较大且活跃度较大的用户进行了详细点的描绘，取情感值小于 20，活跃度大于 0.02 的所有用户，作图 5.36。这是一类在论坛中活动积极、负面情绪较大的用户。这类用户的帖子和评论影响人们对于依法治国的看法。

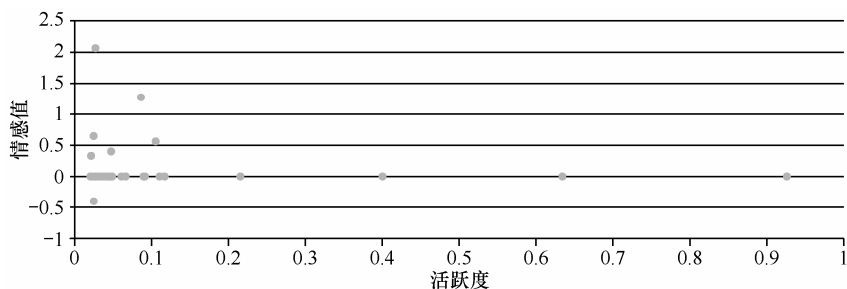


图 5.36 低情感和高活跃度散点图

从图 5.36 可以观察到那些活跃度较高同时情感值较低的用户，论坛管理员要对这些用户进行监察和历史查询，当这些用户出现不正当言论是进行及时处理。

### 5.5.5 专题网络的构建与分析

构建专题网络的方法见节 4.3.5。构建出股市与股票的专题网络，进行可视化，结果见图 5.37。

由图 5.37 可知，“新手 L”和“OK 千里千寻”以及“d61a6ae0d6eb...”三个用户发布的帖子及受众构成的社区较大，选取它们进行分析。

#### 1) 网络指标概述

本节分别从社会网络的整体性指标（表 5.22）、中心性与影响力指标、关联性与“小世界”指标这几个方面进行了分析，以便发现网络中的特殊节点，以便分析个体网络演化与传播。

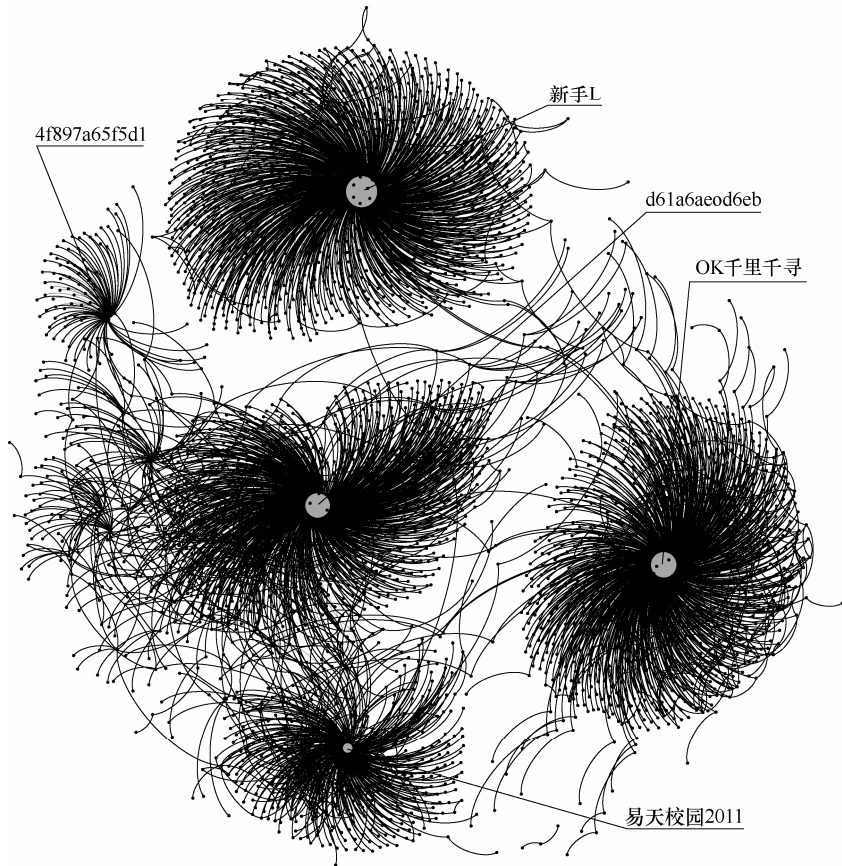


图 5.37 专题传播网络

表 5.22 整体性网络指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
2.470	0.001	9	0.725	4.120	0.617

平均度：较小；图密度：该传播网络图比较稀疏；网络直径：较长，该网络的传播较远；模块化指标：划分出 7 个社区，且划分的质量较高；平均路径长度：长度较大，网络传播较广；平均聚类系数：较小。

2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算，这里取点度中心度排名前 5 的节点，列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名，得到对应的数据如表 5.23 所示。



表 5.23 中心度指标

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	值排名	绝对值	相对值	值排名	绝对值	相对值	值排名
新手 L	3078	0.754	1	974	0.010	68	7060455	0.045	248
OK 千里千寻	1428	0.350	2	0	—	—	7481443	0.043	250
d61a6ae0d6eb...	1109	0.272	3	239355	2.354	1	6085474	0.052	2
易天校园 2011	888	0.218	4	191976	1.888	2	6087508	0.052	25
4f897a65f5d1...	294	0.072	5	0	—	—	5838307	0.055	1

注：—代表不参与相关项。

从表 5.22 中可以看出“OK 千里千寻”“新手 L”等 5 个节点的点度中心度排名前 5 位。从中间中心性来看，“d61a6ae0d6eb...”的中间中心度远远大于其他节点，排在第一位，说明许多节点都需要经过该节点才能建立联系并获取信息。从接近中心性来看，“4f897a65f5d1...”和“d61a6ae0d6eb...”的接近中心度排名前两位，说明这两个节点与其他节点进行交互比较容易。

图 5.37 的点度中心势指标为 0.75%，该指标值很小，表示该图在度数上向某点集中的趋势很不明显；该图的中间中心势指标为 2.35%，该指标值也比较小，表示该图向某些中间节点聚集的程度比较小，即在整个网络中大部分的结点并不需要经过某些中间节点，就能得到信息。

3) 关联性与“小世界”

图 5.37 的特征途径长度为 4.773，凝聚度为 0.008。该图路径长度的具体分布如表 5.24 所示。

表 5.24 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	4063	0.012	5	87793	0.267
2	38599	0.117	6	63042	0.192
3	43458	0.132	7	27890	0.085
4	44232	0.134	8	11980	0.036

图 5.37 的聚类系数为 18.306，节点的聚类系数如表 5.25 所示。

表 5.25 聚类系数指标

用户	聚类系数	点对数
该有的给我	23.167	3
llydxx	0.133	925
易天校园 2011	0.003	55945
笑看风里落花	9.262	21
为之奈何 2014	5.200	15

### 5.5.6 个体网络分析

在按照之前的整体网分析之后，我们从中选取那些节点度数相对较大、传播性比较好的帖子，做了某帖子下参与的所有用户的个体网络。本节选取的是“d61a6ae0d6eb4b7f979691a7a9f6380e”发表的一篇帖子，做可视图（图 5.38），对它进行指标分析，并描述传播路线。

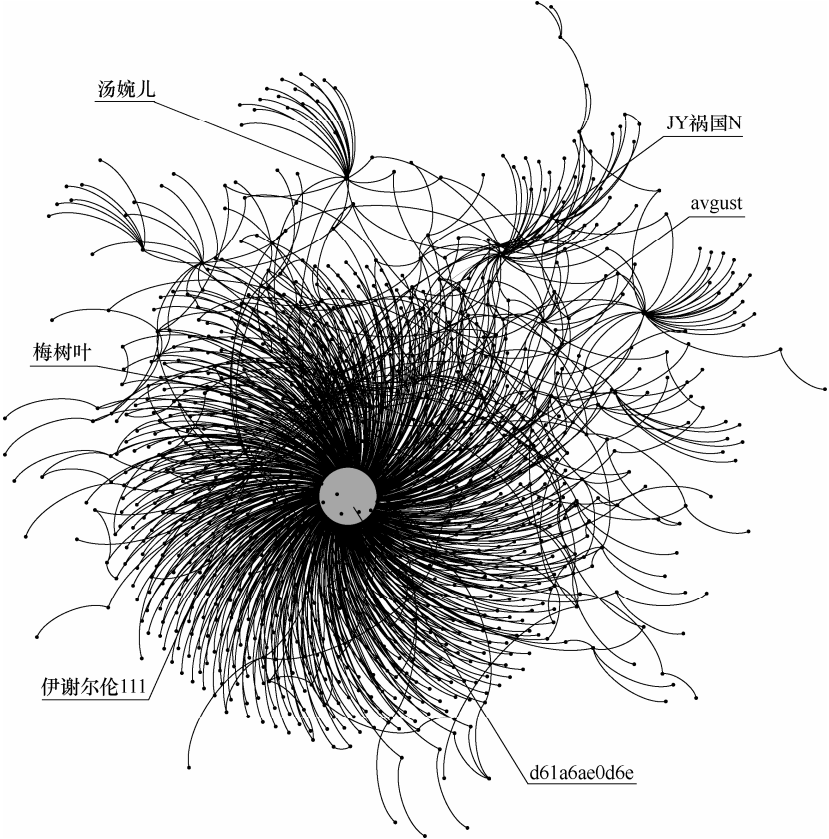


图 5.38 依法治国个体网

从图 5.38 中可以看出，“d61a6ae0d6eb4b7f979691a7a9f6380e”发表的帖子后，大量用户参与的讨论。

#### 1) 结构洞

这里列出节点度数排名前 5 位的用户的结构洞指标，形成相应的数据指标表（表 5.25）。

表 5.25 结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
d61a6ae0d6e...	776	767.594	0.989	0.198	0.956
avgust	14	10.712	0.765	0.383	0.445
梅树叶	18	12.482	0.693	0.573	0.594
汤婉儿	37	24.927	0.674	0.513	0.689
JY 祸国 N	35	33.036	0.944	0.140	0.622

从表 5.25 中可以看出“d61a6ae0d6e...”的节点度数、有效规模和等级度都比较大，说明该居于网络的核心；同时该节点效率最大，限制度最小，表明该节点在网络中的行动最高效，受到网络中其他节点的限制比较小。“梅树叶”对应的效率值较小，限制度较大，说明该节点在网络中受到其他节点的限制比较大，行动力较弱。

2) 诚实中间人

从表 5.26 中可以看出“d61a6ae0d6eb4b7f979691a7a9f6380e”代表的节点作为中介的次数为 18 次，共联络了 153 对关系，其中 151 对关系属于 HBIO 类，151 对关系属于 HBIO1 类，2 对关系属于 HBIO2 类。这说明大部分节点属于纯中介，参与回复主贴的人物之间的关系比较弱。这也表明大部分回复是针对主贴回复的，而只有小部分回复针对的是评论用户评论的。

表 5.26 诚实中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
d61a6ae0d6eb4b7f979691a7a9f6380e	18	153	151	2	0
梅树叶	5	10	10	0	0
汤婉儿	7	21	19	2	0
伊谢尔伦 111	4	6	2	4	0

HBIO 代表纯中介，HBIO1 代表弱中介，HBIO2 代表非中介。

5.6 本章小结

改革开放三十多年以来，中国经济和社会发展取得了举世瞩目的成就，政治、经济、文化等各方面取得了前所未有的进步和发展，但同时也产生了分配

不公、诚信缺失、贪污腐败等问题的滋生和蔓延，引起了广大人民群众의强烈不满。为了更好地适应经济体制改革的步伐，促进市场经济的发展和政治民主化的步伐，党的十八大提出了政治体制改革的议程，目的是在党的领导下和社会主义制度下更好地发展社会生产力，充分发挥社会主义制度优越性。

中国政治体制改革的长期目标是建立高度民主、法治完备、富有效率、充满活力的社会主义政治体制。但是目前，在体制上还存在一些不足，如贪腐现象较多、法制体系不完善等。要解决这些问题，不仅需要进行经济体制改革，还需要进行政治体制的深化改革。

## 第6章 民生领域的相关舆情分析

民生问题是立国之本，改善民生是社会主义的特质和目标。改善民生既是一项重要的政治任务，也是判定科学发展的主要标尺。在改革开放新阶段的发展中，围绕改善民生发展经济文化事业和社会事业，通过推进体制改革与创新为改善民生提供有效的保障，这是科学发展观的内在规定和必然要求，是检验执政效果的实践标准。

我国已跨入全面改善民生的新时代，特别是当前我国的经济的发展、社会发展、政治发展和生态发展之间呈现出结构失衡的状态，即经济发展快，社会发展、生态发展相对缓慢，政治发展相对滞后。这使得我国在发展过程中遇到一些制度、体制和机制方面的问题，最终都以民生问题的形式凸显出来。可见改善民生已成为当代中国发展最核心的政策方向与目标，在改革发展的新阶段，需要在思想、政策、体制等多个层面进行更加深刻的理性思考，通过体制改革和创新为改善民生提供强有力的制度保障。

由于民生问题事关广大人民群众切身利益，因此在十八届三中全会中也采用大量的篇幅来阐释和部署了民生问题的改革，特别强调紧紧围绕保障改善民生这一目标，促进社会公平正义的社会体制改革，改革收入分配制度体系，推进社会各领域的制度创新，以及推进基本公共服务均等化，加快并形成科学有效的社会治理体制，确保社会既充满活力又和谐有序。坚持以人为本，尊重人民主体地位，发挥群众首创精神，紧紧依靠人民推动改革，促进人的全面发展。我们相信，随着我国综合国力的不断增强，民生问题将在更高层次上得以解决，越来越多的群众将能更多更公平地分享经济社会发展的成果。

为了客观地分析民生领域的热点舆情，我们从天涯论坛中选取了2014年度一些民生领域热点专题的帖子进行深入的总结和分析，本章后续部分将从不同的角度展示该领域舆情分析的结果。

### 6.1 民生领域热点专题和热点词分析

本节选取政治领域的相关帖子进行分析，通过对词频进行统计得到该领域的热点词及演化情况，通过对原帖进行文本聚类得到热点事件，从而确定该领域的热点专题。

### 6.1.1 民生领域热点词分析

热点词出现的频率会随着热点事件的发生而快速增长,从热点词的演化过程可以看出不同时间热点事件的主题变化情况。对2014年全年的数据按季度进行词频统计和排名,结果如表6.1所示。

表 6.1 高频词汇统计

2014.01~2014.03		2014.04~2014.06		2014.07~2014.09		2014.10~2014.12	
词语	词频	词语	词频	词语	词频	词语	词频
公务员	38858	公务员	45902	公务员	44568	公务员	46997
社会	39962	社会	42584	社会	42036	社会	33736
服务	37645	服务	40363	服务	39986	服务	32269
部门	29845	部门	31703	部门	31311	部门	25385
国家	28788	国家	30577	国家	30530	国家	24647
保障	27661	保障	29623	保障	29217	保障	23731
工作	25244	工作	26714	工作	26108	工作	21003
管理	15176	管理	16400	管理	16255	管理	13004
群体	12598	群体	13452	群体	13296	群体	10749
孩子	9669	孩子	10648	贪官污吏	10910	辞职	8893
养老	8981	政府	10292	养老	9204	养老	7704
监督	8451	组织	9241	官员	9146	权力	7401
年收入	8337	房子	8971	职业	9083	干部	7328
鼓励	6323	网友	8937	执法	8793	职业	7297
权益	6318	工资	7846	工资	7328	官员	7278
评价	6253	领导	7571	待遇	6803	监督	7270
市场经济	6246	待遇	6920	老百姓	6776	老百姓	5611
发展	5653	权益	6789	翻译	6553	待遇	5559
社保	4474	纳税人	6650	发展	6079	队伍	5366
税务	4122	发展	6205	社保	4756	发展	4767
赡养	4108	环境	4823	省委	4396	发展	4767
企业	3561	社保	4820	老虎	4385	市场	4113
改革	3534	住房	4629	网民	4383	建设	4081
制度	3249	效率	4559	管理者	4376	社保	3928
创业	2522	规划	4526	指挥	4374	补贴	3795
基层	2498	税务	4429	学校	3230	国务院	3730

续表

2014.01~2014.03		2014.04~2014.06		2014.07~2014.09		2014.10~2014.12	
词语	词频	词语	词频	词语	词频	词语	词频
举报	2215	海关	4406	医院	3128	职务	3633
监管	2179	调研员	4406	事故	2499	总理	3629
舆论	2167	医疗	3306	百姓	2498	责任	2118
税收	2166	就业	3035	合同	2460	城市	2111
生育	2152	机构	2591	机会	2459	父母	2077
义务教育	2133	朋友	2569	人才	2454	机关	2076
加薪	2123	父母	2564	属于	2453	基础	2054
公积金	2116	正义	2441	福利	2446	处理	2040
年薪	2107	岗位	2415	西方	2444	调查	2023
婚姻	2090	财产	2405	基层	2440	职责	1826
地震	2089	经费	2377	压力	2411	正义	1825
发言人	2079	老人	2372	竞争	2411	环保	1820
公信力	2062	义务教育	2338	警察	2291	年薪	1816
经济基础	2059	生存	2328	经费	2287	舆论	1815

为了更好地观察词频的演化过程,我们用作图工具 NetDraw 对高频词和时间的对应关系进行了可视化,得到如图 6.1 的高频词演化图。

从热点词和词频来看,有一部分词,如“公务员”“社会”“服务”“部门”“国家”“保障”“工作”“发展”“社保”“管理”“群体”等是全年讨论的热点,这些词基本表明了民生领域的热点话题,尤其是“公务员”一词具有很强的代表性,公务员群体一直是整个社会的舆论焦点,因此全年的讨论词频都位居首位。“养老”“孩子”“父母”“工资”“待遇”“年薪”“基层”“老百姓”等词出现的频率也很高,这些高频词涉及的话题比较广泛,基本涵盖了民生领域“衣食住行”的各个方面。进一步观察可以发现很多热点词的词义相近,比如“工资”“待遇”“职务”“职责”等,同样体现了民生领域的热点话题。

需要说明的是,从统计的角度来看,如果取相当多的词汇来作图,则会发现各个季度中几乎所有的词都出现过。因此,需要对热点词进行筛选,便于能够看到热点事件的变化情况。

### 6.1.2 民生领域热点专题分析

通过对 2014 年民生领域的帖子进行文本聚类得出事件,根据 2.3.2 节的指标进行事件热度的计算,并对这些事件进行摘要的抽取和标题的生成。本节列举热度排名前 10 的热点事件形成表 6.2。

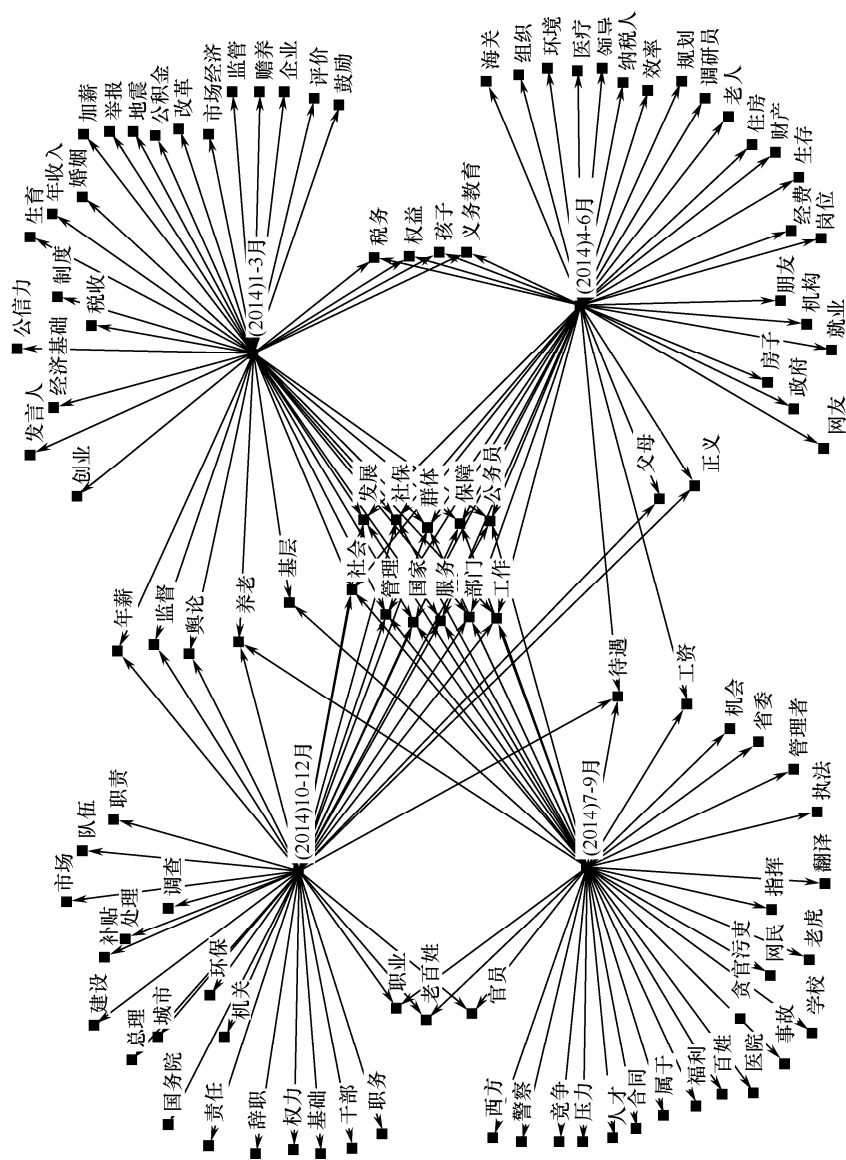


图 6.1 高频词汇演化图



表 6.2 民生领域相关的十大热点专题

热点专题	帖子量/条	总浏览量/条	总回复量/条	参与人数/个	专题热度
教师工资该涨了吗	112	1024032	14738	1972	1.000000
大学生创业历程	96	1057844	993	116	0.989559
谈一谈中国教育现状	223	651358	7261	2720	0.633459
中国物价远高于发达国家	238	487643	9869	2914	0.488543
公务员薪资问题	162	383834	6833	2869	0.382905
养老保险与养老问题现状	192	254505	3356	915	0.249078
延迟退休问题的争议	12	253102	3208	1922	0.248562
事业单位入社保的影响	56	230876	1131	481	0.219510
医疗保障现状	191	154632	2154	917	0.152606
教育公平和资源分配问题	39	21468	128	79	0.021224

从表 6.2 中可以发现,“谈一谈中国教育现状”“教育公平和资源分配问题”“大学生创业历程”等事件分别反映了中国教育与就业的相关问题,“教师工资该涨了吗”“公务员薪资问题”“中国物价远高于发达国家”等事件集中反映了中国网民对特殊群体收入的关注乃至收入分配的问题,“养老保险与养老问题现状”“延迟退休问题的争议”“事业单位入社保的影响”“医疗保障现状”等事件分别反映了社会保障和医疗保障的相关问题。综上所述,教育与就业、收入分配和社会与医疗保障是民生领域中,网民讨论和关注较多的几个热点专题,本报告以教育与就业、收入分配以及社会与医疗保障三个专题为例,对民生领域的网络舆情进行深入的分析。

6.2 民生领域意见领袖分析

在民生领域中涉及的用户包括相关帖子的发布者以及参与浏览和回复这些帖子的舆情受众。通过对网络舆情发布者的关注度分析,可以有效地监测出网络舆情在传播与演化过程中具有影响力的意见领袖。本节采用发布者影响力、发布者活跃度和发布者价值这三个不同维度的指标对舆情发布者进行分析,采用负面指数、参与度和受众影响力对各个专题的舆情受众进行分析。

### 6.2.1 意见领袖

#### 1) 发布者影响力

计算公式见节 2.3.2。公式中参数  $g_1$  取 0.05、 $g_2$  取 0.25、 $g_3$  取 0.3、 $g_4$  取 0.4。计算出的影响力结果如表 6.3 所示。

表 6.3 经济领域论坛用户影响力排名

用户	粉丝数/个	发帖数/条	帖子点击数/次	帖子回复数/条	回帖数/条	影响力
杨胜章 2009	2541	1	14758765	10392	478	0.995055
苦百姓 a	6	1	3836566	42	0	0.651513
说真话好难	519764	1	12093	28	0	0.364228
正风青杨	7749	12	753114	5438	0	0.159950
别了还来	4637	82	713042	6440	2104	0.150364
对世界很迷茫 i	5	1	528040	721	14	0.107051
每天一顿饭	409	2	440022	4990	595	0.091987
sunny9982011	10	1	423652	593	6	0.086018
乔志峰	107823	17	30481	271	0	0.084812
老鹰嘴 3256	108423	5	8065	58	3	0.080612
文侠罗竖一	109110	8	5200	79	0	0.080546
一棵小草 909	14	1	378143	3295	37	0.078325
素秋 R	27	1	364572	7310	79	0.077783
啃咸菜谈天下	54285	4	180638	1150	36	0.076719
荣耀与卑微	608	5	352087	2979	204	0.073332

表 6.3 列出了用户影响力排名前 15 位的用户。“杨胜章 2009”的影响力最高，尽管该用户只发表了一篇与创业相关的帖子，但该帖子引起了网友的广泛兴趣，浏览次数和回复数都排名很高，但其活跃度并不高，这种用户不一定能成为意见领袖；“正风青杨”“别了还来”“荣耀与卑微”等用户发帖数和回复帖子数都比较高，且发表的帖子引起了用户的热烈讨论，这些用户极有可能成为该领域的意见领袖。

#### 2) 发布者活跃度

计算公式见节 2.3.2。公式中参数  $g_1$  取 0.6、 $g_2$  取 0.4，计算结果如表 6.4 所示。

表 6.4 经济领域论坛用户活跃度排名

用户	发帖数/条	回帖数/条	活跃度
0367bb597c5d4793	0	2570	1.000000
9a3ad2572d734795	0	2543	0.989494
别了还来	82	2104	0.866537
8013dad3e49c48b6	0	1637	0.636965
29ec069ff7134ef6b1	0	1632	0.635019
b093a156ba834a23	0	1630	0.634241
坚持到底 53	0	1392	0.541634
南宫羽琦 2014	0	924	0.359533
海外归来客	1	888	0.346109
我不是一棵小小草	1	780	0.304086
至中至正 2011	19	715	0.289300
光明客上访	19	615	0.250389
每天一顿饭	2	595	0.232685
生活收入保障	1	531	0.207198
云裳 2014	2	519	0.203113

表 6.4 列出了发布者活跃度排名前 15 名的用户。这些用户在论坛中较为活跃,同时可以作为重点监视的对象。“别了还来”“至中至正 2011”“云裳 2014”属于活跃性的意见领袖,他们保持着较高的发帖量和回帖量。同时,用户影响力和用户活跃度并不存在严格的正比关系,影响力高的用户活跃度不一定高,有可能是并不活跃的用户发表了一篇大家感兴趣的帖子,在网络上引起了热议,活跃度和影响力都比较高的用户值得重点关注。

### 3) 发布者价值

网络中发布者之间会经常互动,那么发布者的情感价值观也会相互传递、相互影响。发布者的情感倾向成为了舆情监测的重点。

选择排名前 10 的负向情绪用户(表 6.5),可以看出大多数情感倾向为负值的用户,他们可能是舆情监测的重点对象。

表 6.5 发布者情感倾向表

发布者	情感倾向	发布者	情感倾向
海上田园边	-302.4690	hswboy	-12.12650
该起什么名子	-22.74660	现金社会 012	-10.67510
小溪水清又清	-17.33330	yongxu20121225	-9.71000
2c9fc9ed6c634e74843b74a37666367d	-14.60520	网络传销最可耻	-8.79508
ef33be737ee84672b40502178943fa34	-14.06570	中子	-8.39808

### 6.2.2 舆情受众

舆情受众主要指所有被舆情影响到的网络用户，衡量舆情受众主要有三个指标：负面指数，参与度、受众影响力。（计算公式同节 2.3.2）本领域中的三大热点的舆情受众指标如表 6.6 所示。

表 6.6 舆情受众指标

热点专题	负面指数	参与频度	受众影响力
教育与就业	0.43052	0.0166	0.12693
收入分配	0.43242	0.0147	0.15814
社会与医疗保障	0.47689	0.0116	0.14811

## 6.3 专题 1：教育与就业

经过对民生领域的数据进行热点词和聚类分析，选取教育与就业作为民生领域的第一个热点专题。百年大计，教育为本。教育质量的优劣直接关系到民族素质的提高，事关国家经济的振兴。教育成果的经济效用，是要随着劳动能力的社会效用之实现而实现。影响劳动能力社会效用发挥的一个主要条件就是就业。本专题中，可以看到网民对教育公平、中国教育的现状以及大学生就业等问题进行了热烈的讨论和深入的思考。本节将从教育与就业专题的发展简介、专题中网民观点的分类和总结、专题的传播趋势和用户情感以及专题网和个体网分析等方面进行详细叙述。

### 6.3.1 专题发展简介

教育是民族振兴的基石，教育公平是社会公平的重要基础。要全面贯彻党的教育方针，坚持育人为本、德育为先，实施素质教育，提高教育现代化水平，培养德智体美全面发展的社会主义建设者和接班人，办好人民满意的教育。优化教育结构，促进义务教育均衡发展，加快普及高中阶段教育，大力发展职业教育，提高高等教育质量。重视学前教育，关心特殊教育。更新教育观念，深化教学内容方式、考试招生制度、质量评价制度改革，减轻中小学生课业负担，提高学生综合素质。坚持教育公益性质，加大财政对教育投入，规范教育收费，扶持贫困地区、民族地区教育，健全学生资助制度，保障经济困难家庭、进城务工人员子女平等接受义务教育。加强教师队伍建设，重点提高农村教师素质。鼓励和规范社会力量兴办教育。发展远程教育和继续教育，建设全民学习、终身学习的学习型社会。

就业是民生之本。要坚持实施积极的就业政策，加强政府引导，完善市场就业机制，扩大就业规模，改善就业结构。完善支持自主创业、自谋职业政策，加强就业观念教育，使更多劳动者成为创业者。健全面向全体劳动者的职业教育培训制度，加强农村富余劳动力转移就业培训。建立统一规范的人力资源市场，形成城乡劳动者平等就业的制度。完善面向所有困难群众的就业援助制度，及时帮助零就业家庭解决就业困难。积极做好高校毕业生就业工作。规范和协调劳动关系，完善和落实国家对农民工的政策，依法维护劳动者权益。

教育与就业是民生领域最为基本的一项，本节将对民生领域中教育与就业相关的舆情进行分析。表 6.7 是该专题下的一些热点帖子以及相关统计指标。

表 6.7 教育与就业专题热帖列表

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
周有光老先生说，民国时期的教育是最好的	sweetswing	2014/3/31	113126	1918	555	337	57.1170
【天涯头条】我是一名中学教师，痛揭中学教育之病态！	荣耀与卑微	2014/6/30	328169	1778	610	379	53.9323
翻身，教育不公，维权不止！（让图片说话）	云裳 2014	2014/6/4	32217	1326	721	274	63.7702
感谢邓小平恢复了高考！受过高等教育的顶吧！	在黑暗中等 待光明	2014/2/2	24024	851	227	171	41.8792
近来看新闻，发现淡化支持高校毕业生就业，而着重自主创业	kumsing	2014/5/14	14894	723	126	103	24.1120
留美多年，中国教育水平已世界领先，当心精公的“素质教育”坑民族	啄木鸟老师	2014/10/27	232054	737	297	116	11.6336
我是 80 后乡村女教师，我也说说现在的教育现象	奕彤宝贝	2014/12/11	58347	315	129	68	5.72583
【风青杨时评】727 万大学生毕业为何就业难？	正风青杨	2014/5/29	94992	307	97	85	62.0110

### 6.3.2 观点分类与观点列表

本节以论坛中网民的回复为研究对象，通过观点挖掘和语义分析形成本专题内的核心观点，并对其进行分类和详细描述。

#### 1) 观点分类

根据观点涉及的特征词汇和领域以及网民发表的各个观点内容本身之间存在的相关性与差异性，可以将所有评论分为以下三个主题，每个主题对应的观点人数比例图为图 6.2。

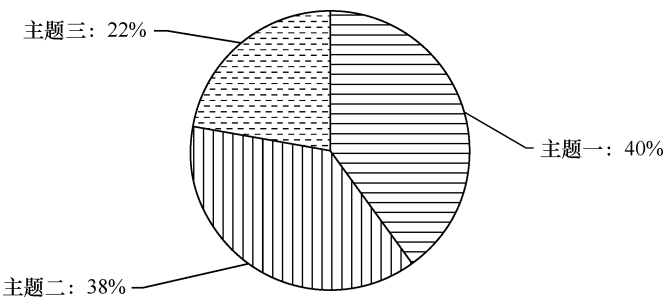


图 6.2 “教育与就业”专题观点分类比例图

#### 主题一：教育公平问题，优质教育资源稀缺

此分类观点包含四个关键特征词：公平、资源分配、招生政策、入学。对应的观点包括：

观点 1. 释放教育公平，还孩子就近入学的权利。

观点 2. 异地高考迈出教育公平新一步，教育水平地区差异大。

观点 3. 教育资源的紧张导致民众合理诉求的产生，优质教育资源不足。

观点 4. 教育公平是实理中国梦的根基。应该深化考试招生制度改革，促进教育公平。

#### 主题二：谈一谈中国教育的现状和未来

此分类观点包含五个关键特征词：教育、现状、应试、改革、实践。对应的观点包括：

观点 5. 中国教育是失败的，教育出来的学生没有创新能力，英语不会说，只能用来应付考试。应该加强实践教育。

观点 6. 中国的教育体制存在弊端，需要进行体制改革。

观点 7. 中国目前的高考招生是相对公平的，只是这个独木桥留给普通民众的空间越来越窄了。

观点 8. 精神源自文化，文化来自教育。中国教育应该重视传统文化教育。

观点 9. 高考改革不应该只关注“考什么”，更应该关注怎么考。

### 主题三：大学生该就业还是该创业

此分类观点包含四个关键特征词：就业、创业、大学生、选择。对应的观点包括：

观点 10. 大学生上大学就是应该就业，创业有风险，应该先就业积累经验。

观点 11. 引领大学生创业正当其时。创业虽然难，会有很多问题，但是能够实现自我价值，实现抱负。

观点 12. 大学生创业存在诸多挑战，不符合市场经济规律。

## 2) 观点分类详细分析

为了更好地从宏观向微观的深入分析，本节对每个主题内的网络用户分布以及网络民意进行深入分析，希望通过网络用户真实的反馈信息来反映他们对这些问题的思考与建议。

### 主题一：教育公平问题，优质教育资源稀缺

我们对网民在这个主题内的观点进行进一步挖掘和梳理，发现存在 5 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 16.23%的网友认为应该释放教育公平，还孩子就近入学的权利。

例如，网友“Mr、独木”说：“就近入学，是教育公平最基本的体现，一些教育专家曾提出，就近入学最能体现教育公平，就近入学是目前大家能够认识到的解决教育公平问题最为现实的方法。而深圳市豪方片区所显示出的不公平教育问题，明显地反映出当地市政府以及教育部门对公平教育方针的不彻底贯彻与执行。教育问题，不仅仅只是教育本身，更是社会问题。从这个角度出发，政府要从社会整体发展的角度来看待教育改革。一直以来，教育问题都备受家长和社会的关注。关于教育公平问题，不能仅仅停留在表面宣传，更应该从实际出发，切实地听听家长们对于孩子教育问题的意见。为了所有孩子的教育，请呼吁教育公平合理！”

观点 2. 26.50%的网友认为教育水平地区差异大，异地高考是迈开教育公平新一步。

例如,网友“东营 101 网校”说:“教育资源均衡配置是更深入、更全面也更艰难的变革,必须经历积跬步而致千里的渐变过程。教育改革如果仅仅停留在“异地高考”层面,还远远不够。大城市公共资源容量有限,完全不设门槛、单边放开异地高考,可能导致流动人口大量涌入,挑战城市管理,引发群体矛盾。对教育部以及各地政府而言,夯实并彻底改善农村、特别是中西部农村的基础教育和高中教育,比起异地高考,这样的改革更深入、更全面,当然也更艰难。”

观点 3. 34.17%的网友认为教育资源的紧张导致民众合理诉求的产生,优质教育资源不足。

例如,网友“sp 暴君”说:“刘强东说的现实很严峻。不是我们的高考这个考试制度不公平,而是高考背后的教育资源出现了严重的两极分化。在这方面,中国做得比美国要好,我们的高考制度比美国的录取制度更有利于穷人的孩子获得更好的教育。但是还远远不够,需要改进的地方还很多。”

观点 4. 24.10%的网友认为教育公平是实理中国梦的根基。深化考试招生制度改革,促进教育公平。

例如,网友“CY7871360”说:“教育公平既是和谐社会的重要内容,又是和谐社会的重要基础,有了教育机会的公平、教育过程的公平和教育质量的公平,社会弱势群体才有可能与社会其他阶层在同一起跑线上起跑,才有通过知识改变命运的可能性,社会各阶层才有正常流动、分化的可能,全社会才能充满活力、安定有序,而不至于让社会弱势群体无路可走,铤而走险。因此,不断探索教育改革和实践,努力实现教育公平,保障受教育者的权益,是构建和谐社会的基本途径。”

综上,在本主题中,具有相应子观点的评论分布比例如图 6.3 所示。

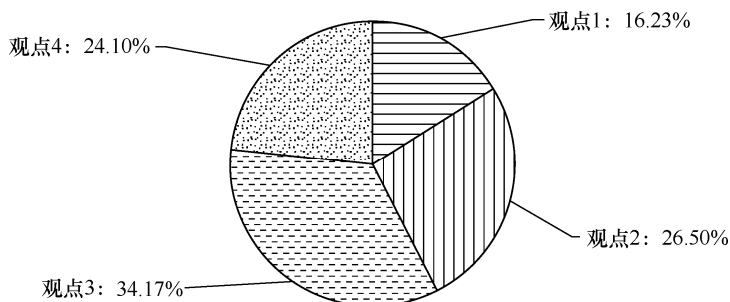


图 6.3 主题一细分子观点比例图



## 主题二：谈一谈中国教育的现状

我们对网民在这个主题内的观点进行进一步挖掘和梳理，发现存在 5 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 20.10%的网友认为中国教育是失败的，教育出来的学生没有创新能力，英语不会说，只能用来应付考试。应该加强实践教育。

例如，网友“ygytopbest”说：“中国教育有两大无知：一是把实践等同于文盲，认为实践就是丢读书人的面子；二是把应用等同于技校，认为大学教应用就降低了办学层次。”

观点 2. 29.90%的网友认为中国的教育体制存在弊端，需要进行体制改革。

例如，网友“她的肩膀”说：“中国在 1997 年提出素质教育，倡导把学生带出课堂的教育方针，可是根据近年来的表现，中国学生真的走出课堂了吗？中国需要的，是能够把想法和知识真正融为一体并且付诸行动的那些人，也只有这些人，才称得上真正的人才，现在这社会，缺少的是技术，不是文凭！教育改革，已成为中国迈向世界强国的最大绊脚石，至于怎样挪动这块大石，就需要全体人民和广大学生团体的共同努力了，中国教育不能再这样下去了，中国教育必须改革，而且必须从头改革。”

观点 3. 11.67%的网友认为中国目前的高考招生是相对公平的，只是这个独木桥留给普通民众的空间越来越窄了。

例如，网友“CY7871360”说：“教育公平既是和谐社会的重要内容，又是和谐社会的重要基础，有了教育机会的公平、教育过程的公平和教育质量的公平，社会弱势群体才有可能与社会其他阶层在同一起跑线上起跑，才有通过知识改变命运的可能性，社会各阶层才有正常流动、分化的可能，全社会才能充满活力、安定有序，而不至于让社会弱势群体无路可走，铤而走险。因此，不断探索教育改革和实践，努力实现教育公平，保障受教育者的权益，是构建和谐社会的基本途径。”

观点 4. 12.34%的网友认为精神源自文化，文化来自教育。中国教育应该重视传统文化教育。

例如，网友“至中至正 2011”说：“学校作为中华文化的主载体、主平台，本来是天经地义，毋庸置疑的。可是目前中国学校教育在这方面欠缺极其明显，中华文化没有放到应有位置，缺乏足够的分量。接受完中国教育，人民很难感受到中华文明的灿烂辉煌，很难了解中华文化的精神，只会把中华文化当做一种古董纪念，没有把中华文化作为一个活着的生命体。当前，加强对学生进行

中华文化教育应该成为一项重要而紧迫的战略任务。”

观点 5. 26.80%的网友认为高考改革不应该只关注“考什么”，更应该关注怎么考。

例如，网友“金永花”说：“关于高考英语的废存问题，社会各界都从不同角度展开了激烈辩论。据了解，高考总体改革方案将于 2014 年上半年出台。从之前披露过的各种改革方案看，此次高考改革基本围绕‘分不分文理科’和‘提高语文比例，降低或取消英语’两点展开。然而，在笔者看来，这些改革方案都把注意力集中在了‘考什么’上，而对我国现行高考方案的最大问题‘怎么考’却都没有触及。”

综上，在本主题中，具有相应子观点的评论分布比例如图 6.4 所示。

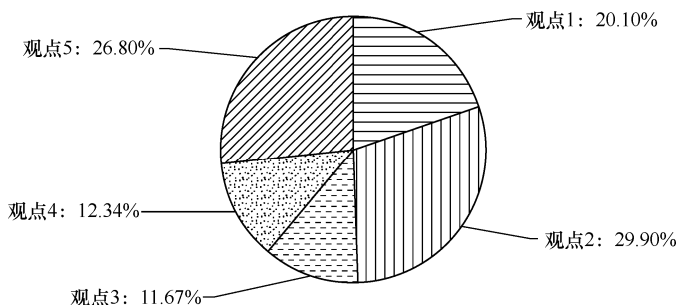


图 6.4 主题二细分子观点比例图

### 主题三：大学生该就业还是该创业

我们对网民在这个主题内的观点进行进一步挖掘和梳理，发现存在 3 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 26.30%的网友认为大学生上大学就是应该就业，创业有风险，应该先就业积累经验。

例如，网友“古月弦 2”说：“大学生有创业热情，但由于经验欠缺、能力不足、意识偏差等原因，导致创业成功率明显偏低。”

观点 2. 31.20%的网友认为引领大学生创业正当其时。创业虽然难，会有很多问题，但是能够实现自我价值，实现抱负。

例如，网友“奋斗的小鸟 12123”说：“让青年的活力和创造力充分迸发，市场的作用是决定性的，设立专项风险投资基金、建设专用股权交易平台、提供专业咨询服务……我们期待创业之花开遍校园，让每位今天的大学生，都有条件有能力成为明天的创业达人。”

观点 3. 42.50%的网友认为大学生创业存在诸多挑战, 不符合市场经济规律。

例如, 网友“目翰勉自”说: “潜台词创业遵循市场经济规律, 政府不能过多指导, 主张先就业了解社会、了解市场, 有能力者自然创业; 事实创业成功概率很低, 不是所有人适合创业, 市场规律是公平的, 市场竞争是残酷的, 风险非常大, 政府只提供创业条件, 媒体宣传创业成功事例, 这样只能增加大学生创业浮躁心理; 政府应遵循市场经济规律, 让有能力者经过就业洗礼、有风险承受能力的自然创业。否则给不谙世事的毕业大学生盲目创业失败造成心理伤害。”

综上, 在本主题中, 具有相应子观点的评论分布比例如图 6.5 所示。

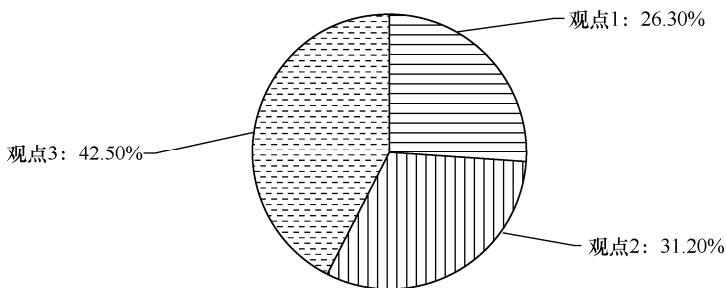


图 6.5 主题三细分子观点比例图

### 3) 观点总结

科技教育和人才是国家发展的基础和根本, 必须始终放在重要的战略位置。当前的中国教育与就业领域出现了很多值得注意的问题。首先是教育公平的问题, 目前的教育资源分配不均衡, 地区教育发展不均衡, 优质教育资源的稀缺导致了很多现实社会问题, 如适龄孩子入学难、农村地区教育资源不足、异地高考争议和中考高考加分政策标准的争议等, 这些问题严重阻碍了教育公平的实现, 引起了社会舆论的广泛关注。其次是广大网友对教育现状的讨论, 大部分网友认为教育现状令人不甚满意, 需要深刻反思, 当前的教育政策也需要进一步调整。最后是大学生就业和创业的问题, 就业难是一个世界性的难题, 尤其在那些人口多、经济尚不发达的国家里更为突出, 即使在发达的资本主义国家里也仍有就业难的现象存在。教育的发展从总体上来说, 对推动就业有积极作用, 但也会产生一定的压力。创业是解决就业问

题的另一个途径,一方面能够实现自就业,另一方还能扩大就业。但是创业可能会遇到很多问题,需要审慎面对。

### 6.3.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量,专题情感分析的基础是帖子回复内容的情感趋向,涉及情感计算。

#### 1) 影响力

专题的传播过程中流量是指相关帖子的回复数量,发展峰值出现在舆论讨论最激烈的时间点。很明显,流量与影响力成正比关系,即流量越大影响力也就越大。因此,本节分别对新增回复数和总回复数按照时间进行统计,然后对统计结果进行可视化得到波线图(图 6.6,图 6.7),以便发现专题传播的发展峰值。

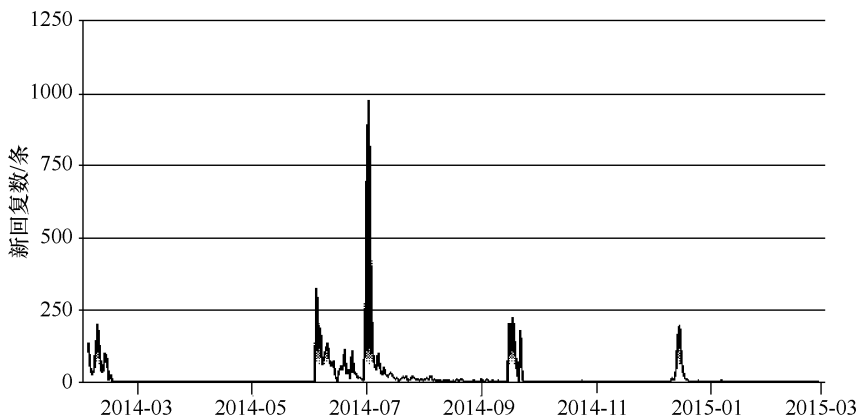


图 6.6 专题日流量时间趋势

图 6.6 中,专题的日流量峰值出现在 2014 年 7 月 2 日。2014 年 6 月 30 日网友“荣耀与卑微”发表了名为“我是一名中学教师,痛揭中学教育之病态!”的帖子,一经发表,该帖子便引起了网友的广泛热议,帖子发表后的第三天,达到峰值,当天评论量达到 976 条。并且该帖子热度持续了一周以上,从专题总流量趋势图(图 6.7)中,可以明显看到七月初帖子回复数骤增的过程。

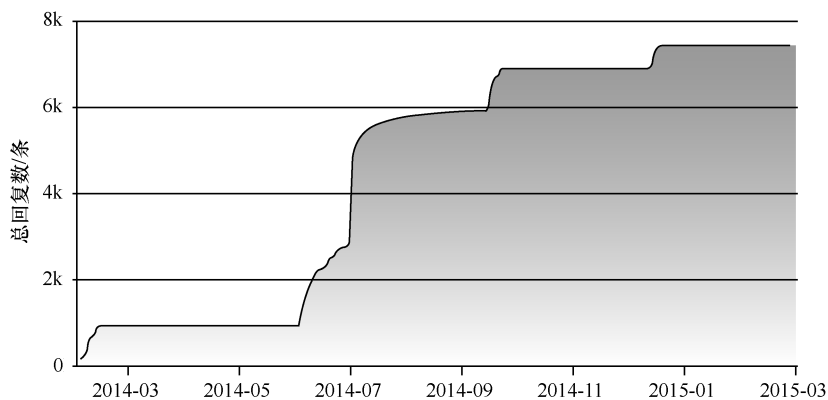


图 6.7 专题总流量时间趋势图

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库，对专题内的回复内容进行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图（图 6.8）和专题情感分类图（图 6.9）。

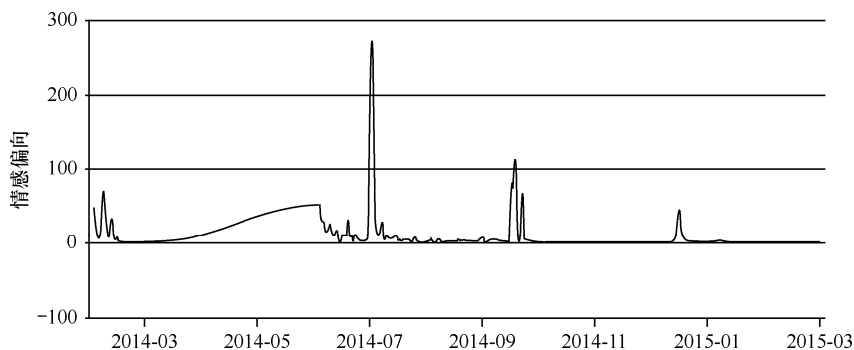


图 6.8 专题情感趋势图

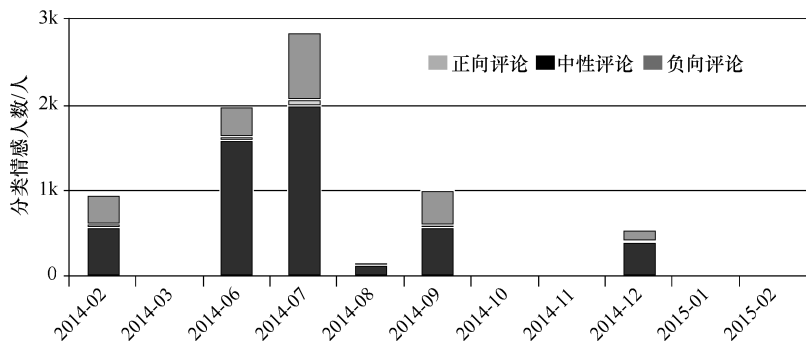


图 6.9 专题情感分类图

图 6.8 与图 6.9 中, 在 2014 年 2 月、6~9 月和 12 月这六个月里, 随着评论数量的增加, 评论的情感趋势也比较高, 6 月总评论数接近 2000 条, 7 月总评论数接近 3000 条, 其他月份的总评论数都不超过 1000 条。明显可以看到, 负向评论的远大于正向评论和中性评论的数量, 2014 年 6 月, 负向评论占总评论的比例达到了 80%, 7 月是 70%, 一直维持在较高的比例, 这说明大部分网民的情绪以愤懑、牢骚、抱怨、起哄为主, 情感倾向比较明显, 将网络社区作为自己情绪宣泄的平台, 这部分用户的盲从性比较大, 很容易受到有心人士的挑唆和煽动, 只有一部分人是就事论事、客观发表自己的看法, 这部分人已成为论坛社区的意见领袖或舆情引导者。

### 6.3.4 专题用户情感的分析

随着近十年来论坛的兴起, 网民数量也急剧上升, 网民作为网络舆情的一个非常重要的因素, 其行为和言论会决定了整个舆论的倾向。而在网民中那些发言多、回复多的“意见领袖”或者“领军人物”往往会主导整个舆论的走向。所以说, 探索专题下的人物情感可以帮助我们发现该专题下的情感正向用户, 同时也可以帮助论坛找出到处散发消极影响的网民, 剔除该用户来净化论坛环境。

本节对各个专题下的人物情感进行了统计分析, 由于活跃度对情感的影响较大, 所以本节构建了分别以情感值和活跃度为两个坐标轴的专题下任务的散点图 (图 6.10)。

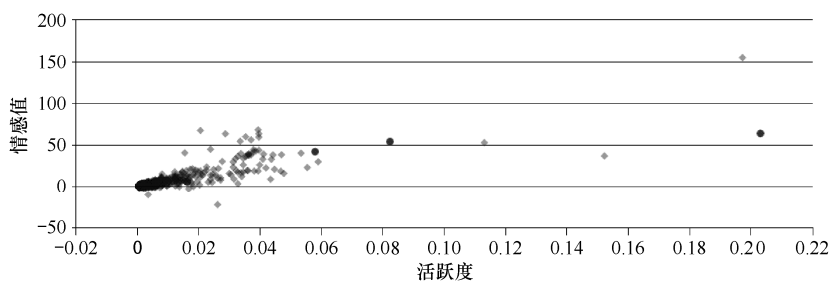


图 6.10 基于活跃度和情感值的用户散点图

圆点代表发帖人, 方点代表回帖人

从图 6.10 可以看出, 本专题中发帖人的情感值和活跃度都相对比较高, 因此可以将这几个人视为该专题的意见领袖。绝大部分回帖人都集中在情感值和活跃度较低的区域, 这部分人大多数时候只是看到事情的表面, 他们的情绪很容易受到所谓“主流声音”的影响, 同时也容易受人煽动和利用。

为了进一步进行分析, 本节将那些情感值负性较大并且活跃度相对较大的人进行了详细点的描绘, 取情感值小于 20, 活跃度大于 0.02 的所有用户, 得图 6.11。

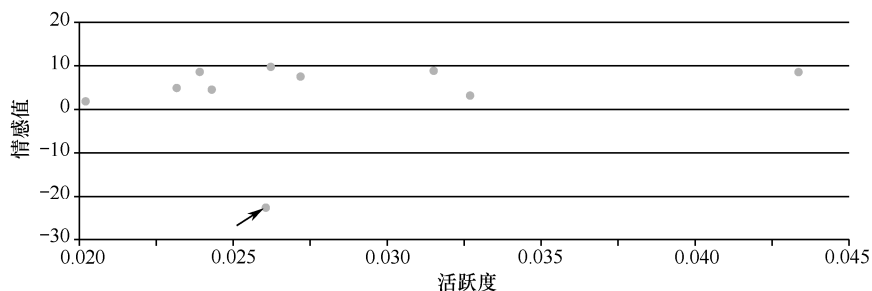


图 6.11 低情感和高活跃度散点图

图 6.11 中箭头所指向的点在论坛中较活跃，且情感值比较负面，该用户可能是宣扬消极情绪的用户，论坛管理员可以对该用户进行监控，以便及时控制网站中的负面情绪的传播，营造健康的网络环境。

### 6.3.5 专题网络的构建与分析

选取教育与就业专题下关于教育公平与大学生就业的相关帖子及有效评论构建专题网络。这里选取的帖子分别为“荣耀与卑微”“云裳 2014”“在黑暗中等待光明”“奕彤宝贝”“正风青杨”发表的有关教育和大学生就业现状的帖子。从这五篇帖子的相关评论中抽取评论关系，构建专题网络，进行可视化结果如图 6.12 所示。

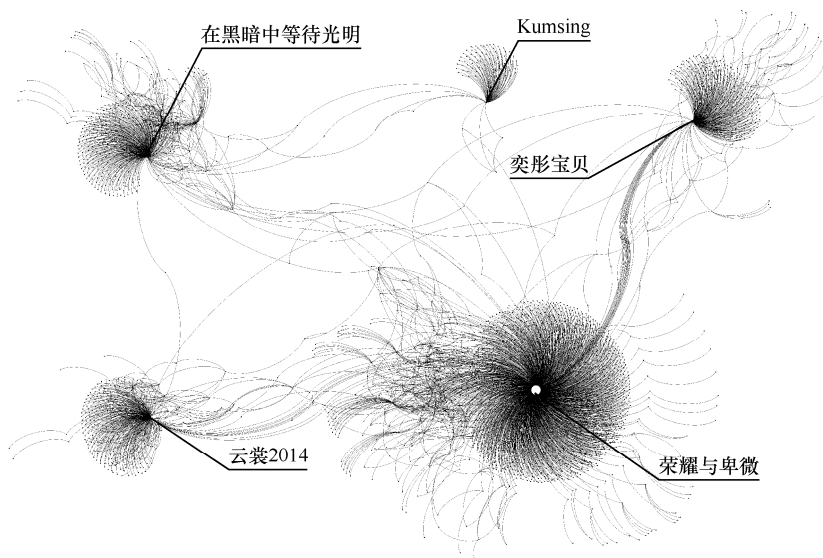


图 6.12 专题传播网络

### 1) 网络指标概述

这里列举了该专题传播网络的一些基本网络指标（表 6.8），使读者对该网络有一个初步的认识。该网络共由 2863 个节点和 4002 条边构成。

表 6.8 基本网络指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
1.398	0.001	12	0.736	4.514	1.301

平均度：较小，平均加权重为 2.788；图密度：该传播网络图比较稀疏；网络直径：较长，该网络的传播较远；模块化指标：划分出 7 个社区，且划分的质量较高；平均路径长度：长度较大，网络传播较广；平均聚类系数：较小。

### 2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算，这里取点度中心度排名前 8 的节点，列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名，得到对应的数据如表 6.9 所示。

表 6.9 中心度指标

用户	点度中心度			中间中心性			接近中心性		
	绝对值	相对值	值排名	绝对值	相对值	值排名	绝对值	相对值	值排名
荣耀与卑微	2074	57.862	1	911162	85.855	1	1341070	55.166	91
云裳 2014	1657	10.342	2	418666	18.877	2	1345491	34.333	175
Kumsing	949	3.040	3	0	5.691	2782	7944912	29.270	433
在黑暗中等 待光明	774	11.740	4	242678	24.391	3	254937	35.797	1
walmm2012	507	0.245	5	4788	0.000	54	1347885	25.576	247
奕彤宝贝	386	10.273	6	176538	19.496	6	1343728	34.407	171
东方红日云 飞扬	101	1.992	7	64477	1.309	13	257632	26.840	2
中将杨立仁	93	0.769	8	33	0.038	226	257668	26.485	3

从表 6.9 中可以看出“荣耀与卑微”的中间中心度远远大于其他节点，排在第一位，说明许多节点都需要经过该节点才能建立联系并获取信息。“云裳 2014”和“在黑暗中等待光明”的中间中心度也相对比较大，这两个节点也各自控制了一部分资源。“在黑暗中等待光明”“东方红日飞扬”“中将杨立仁”的接近中心度依次位列前三位，说明这几个节点与其他节点进行交互比较容易。

图 6.12 的点度中心势指标为 0.15%，该指标值很小，表示该图在度数上向某点集中的趋势很不明显；该图的中间中心势指标为 11.12%，该指标值较大，



表示该图具有中间节点的程度比较大，即在整个网络中有相当部分的结点需要经过某些中间节点，才能得到信息。

3) 关联性与“小世界”

图 6.12 的特征途径长度为 4.514，凝聚度为 0.035。该图路径长度的具体分布如表 6.10 所示。

表 6.10 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	3957	0.004	7	64717	0.061
2	181115	0.169	8	27541	0.026
3	146289	0.137	9	18551	0.017
4	152826	0.143	10	7414	0.007
5	310950	0.291	11	1094	0.001
6	154935	0.145	12	52	0.000

图 6.12 的聚类系数为 1.301，加权的聚类系数为 0.004。节点聚类系数如表 6.11 所示。

表 6.11 节点聚类系数

用户	聚类系数	点对数
荣耀与卑微	0.000	370340
云裳 2014	0.002	43660
kumsing	0.000	3741
在黑暗中等待光明	0.003	56280
奕彤宝贝	0.000	43071

6.3.6 个体网络分析

在按照之前的整体网分析之后，我们从中选取那些节点度数相对较大、传播性比较好的帖子，做了某帖子下参与的所有用户的个体网络。本节选取的是用户“荣耀与卑微”发表的一篇帖子，做可视图（图 6.13），对它进行指标分析，并描述传播路线。

1) 结构洞

这里列举节点度数排名前 5 位的用户的结构洞指标，形成相应的数据指标表（表 6.12）。

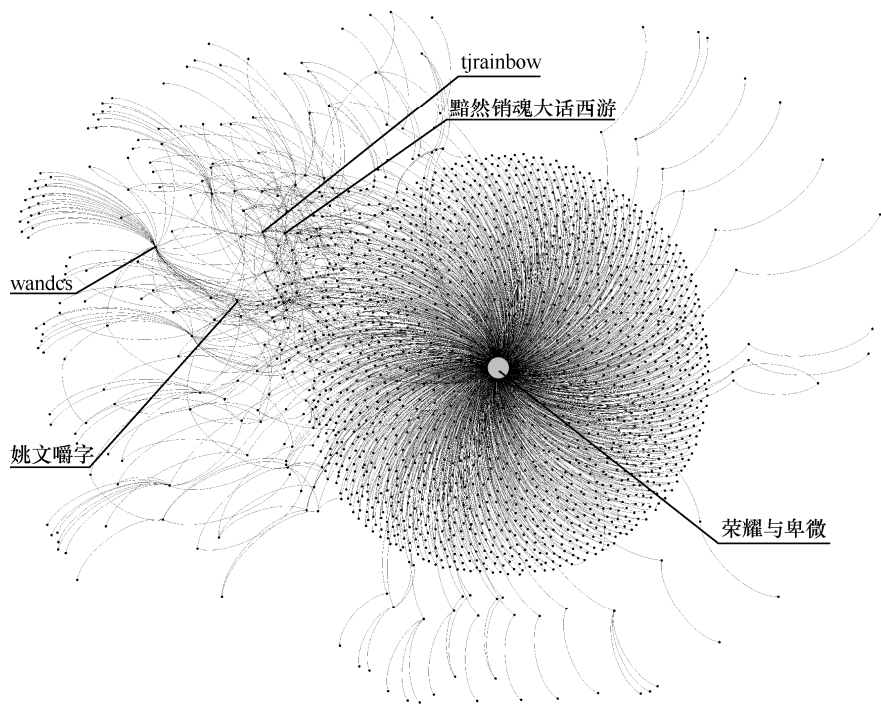


图 6.13 教育与就业专题个体网络

表 6.12 节点结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
荣耀与卑微	1657	1619.376	0.977	0.228	0.989
wandcs	49	48.092	0.981	0.093	0.575
tjrainbow	22	19.021	0.865	0.318	0.602
黯然销魂大话西游	21	16.267	0.775	0.460	0.645
姚文嚼字	17	15.105	0.889	0.290	0.435

从表 6.12 中可以看出“荣耀与卑微”的节点度数、有效规模和等级度值均为最大，其余用户则远远小于该用户，这说明“荣耀与卑微”居于网络最为核心的位置。“wandcs”的效率最大，限制度最小，表明该节点在网络中的行动最高效，受到网络中其他节点的限制比较小。“黯然销魂大话西游”对应的效率值较小，限制度较大，说明该节点在网络中受到其他节点的限制比较大，行动力较弱。“黯然销魂大话西游”“tjrainbow”“wandcs”等节点的等级度比“荣耀与卑微”的值小，又远大于其他节点，位于网络局部中心的位置。

## 2) 诚实中间人指数

这里列出部分代表性用户的诚实中间人指标，形成相应的数据指标表如表 6.13 所示。

表 6.13 诚实中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
荣耀与卑微	77	2926	84	19	7
黯然销魂大话西游	6	15	9	2	4
Tjrainbow	5	10	7	2	1
像疯一样战斗	4	6	3	1	2
xiatn	4	6	4	1	1

HBIO 代表纯中介，HBIO1 代表弱中介，HBIO2 代表非中介。

从表 6.13 中可以看出“荣耀与卑微”代表的节点共联络了 2926 对关系，其中 84 对关系属于 HBIO 类，19 对关系属于 HBIO1 类，7 对关系属于 HBIO2 类。“黯然销魂大话西游”“tjrainbow”等节点也连接了少量节点对，在其中充当了中介。可以看出，“荣耀与卑微”发表的帖子后，大量用户参与的讨论。在回复的用户中，“黯然销魂大话西游”“tjrainbow”等用户发表的评论引起了部分网民的局部讨论，影响比较小。在该网络的传播过程中，没能引起主题的进一步扩散和演化，用户的讨论主要是针对原帖内容的。

## 6.4 专题 2：收入分配

经过对民生领域的数据进行热点词和聚类分析，选取收入分配问题作为民生领域的第二个热点专题。社会公平是构建和谐社会的一项重要内涵。合理的收入分配制度是社会公平的重要体现，是与人民群众切身利益息息相关的问题。党的十七大报告指出：要坚持和完善按劳分配为主体，多种分配方式并存的分配制度，健全劳动、资本、技术、管理等生产要素按贡献参与分配的制度，初次分配和再分配都要处理好效率和公平的关系，再分配要更加注重公平。逐步提高居民收入在国民收入分配中的比重，提高劳动报酬在初次分配中的比重。着力提高低收入者收入，逐步提高扶贫标准和最低工资标准，建立企业职工工资正常增长机制和支付保障机制。这些制度政策的最终目的都是要在兼顾效率的同时保障社会公平，提高居民收入，保障人民群众的切身利益。本专题中，可以看到网民对公务员薪资、教师待遇以及高消费和低收入等问题进行了热烈的讨论和深入的思考。本节将对该专题下的热点事件进行进一步分析，以了解网民对相关问题的讨论、思考与见解。

### 6.4.1 专题发展简介

表 6.14 是收入分配专题中的几大热帖。在专题信息的传播过程中,大量的网民对原始帖子进行了回复,其中产生了一些新颖的且具有一定影响力的观点,这些观点大多是发帖者观点的扩展和延伸,进一步引发了网民的关注和讨论。

表 6.14 收入分配专题热帖列表

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
湖北省全体省编教师致教育厅的一封信	素秋 R	2014/9/24	364572	4223	2229	858	2.55000
【桃花乱开】为什么 - 中国物价为什么远远高于美国等发达国家	看那桃花朵朵开	2014/5/3	170489	1847	555	365	11.49890
【忠言时评】公务员退休金高于老百姓才是公平吗?	忠言	2014/1/8	90827	1098	376	316	9.72159
我也谈一谈“公务员”的工作和收入(长文)	elpasa	2014/1/23	94950	1176	311	247	6.94602
禁令让隐性收入没了多少公务员要扯乎	直隶巴人	2014/1/10	151801	961	290	273	0.78200
调查贴: 你是自干五还是精英, 你的学历是什么, 你的年龄多大, 你的收入	豆豆躲猫猫	2014/12/10	39489	754	213	257	37.11030
惊人啊! 美国人均收入是中国的 34 倍, 物价却是中国的五分之一!!	正红旗旗主六	2014/9/26	75449	547	311	163	3.53236
统计局公布 2013 中国人均年收入 45676 元。我只想说说“呵呵”	sdhl2001	2014/5/27	37513	526	191	144	-0.50000
【风青杨专栏】嫌工资低公务员为何不辞职?	正风青杨	2014/11/20	144815	1821	558	668	62.01100

### 6.4.2 观点分类与观点列表

根据网络数据信息,经过对网民在回复过程中的观点挖掘与倾向进行语义的处理,形成了本专题内的核心观点。

#### 1) 观点分类

根据观点涉及的特征词汇和领域以及网民发表的各个观点内容本身之间存在的相关性与差异性,可以将所有评论分为三个主题,经过文本清洗、语义分析和观点聚类这三个核心步骤,将提取到的观点分到相应主题。各个主题的观点比例图如图 6.14 所示。

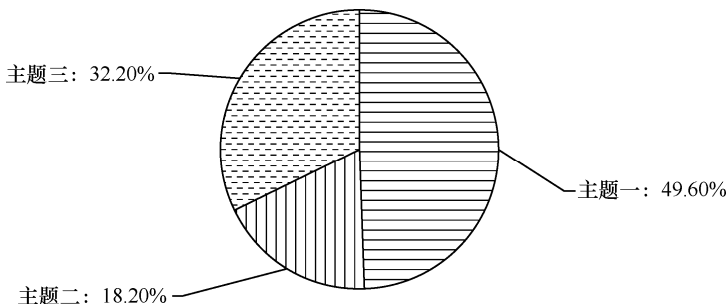


图 6.14 “收入分配”专题观点分类比例图

#### 主题一：社会舆论对公务员薪酬是否该涨的讨论

此分类观点包含三个关键特征词：公务员、薪酬、体制改革。对应的观点包括：

观点 1. 政府机构过于冗余，部分公务员工作清闲，是人们的理想工作，不应该涨工资。

观点 2. 公务员也要养家，薪酬过低，应该加薪。

观点 3. 大多数公务员安于现状，没有抱负消极怠工，脱离群众。

观点 4. 政府机关必须精简机构，加强公务员队伍建设，真正造福人民，在此基础上可以涨工资，不能无差别涨工资，助长不良风气。

#### 主题二：社会舆论对教师待遇问题的讨论

此分类观点包含三个关键特征词：教师、待遇、公平。对应的观点包括：

观点 5. 省招老师和县招老师同工不同酬，省招老师的待遇多年一直未变。

观点 6. 如果湖北省教育厅不能处理好这次省招教师的合理诉求，必然会

极大地打击年轻教师的积极性。

观点 7. 改变某一教师群体(如贫困山区的教师)待遇的现状是非常难的, 不一定会得到重视, 要靠自己的努力向上走。

观点 8. 教师的待遇确实比较差, 给教师提高待遇能帮助教师全心投入工作。

### 主题三: 中国高物价高消费的现状和国民收入的矛盾

此分类观点包含三个关键特征词: 物价、消费、上涨。对应的观点包括:

观点 9. 中国进入了货币贬值、物价飞涨的时代。

观点 10. 物价飞涨, 工资的涨幅远远低于物价, 几乎入不敷出, 在大城市生活压力巨大。

观点 11. 中国的物价水平高于美国澳大利亚, 民众生活困难。

观点 12. 认为美国物价低是非常无知的, 中国的物价没有美国高。

观点 13. 公务员工资是抵挡物价上涨的第一道防线。

## 2) 观点分类详细分析

为了更好地对相应的观点分类进行深入的分析, 本节对每个分类观点内的网络用户分布以及网络民意进行了分类处理, 希望通过网络用户真实反馈的信息, 来反映他们对这一些问题的思考与建议。

### 主题一: 社会舆论对公务员薪酬是否该涨的讨论

在进一步的数据分析与研究中, 对网民针对“公务员薪酬是否该涨”主题内的深入观点进行梳理与分析, 发现还存在 4 个细分的子观点。这些细分的子观点所反映的内容如下:

观点 1. 22.99%的网友认为政府机构过于冗余, 部分公务员工作清闲, 是人们的理想工作, 不应该涨工资。

例如, 网友“chenxuebin914”说: “不应该涨工资。目前中国的公务员分两类人: 一类人, 高学历, 高智商, 正常途径考取公务员。这类人更不应该涨工资, 看看欧美国家, 这类人才做公务员会不会太浪费? 这类人才不做公务员而到企业做高管, 或创业, 得到的福利待遇会比公务员少? 创造的价值会比公务员少? 二类人, 富二代, 官二代, 在中国很多现有公务员有多少是动用关系才进去的? 大家想想! 这类人应该加工资吗? 既然公务员工资低福利差, 干嘛还要花钱找关系进公务员? 在社会上做点什么不好? 既然公务员工资低, 福利差, 干嘛还那么多人报考? 大家都不去报名, 不想着去当官, 公务员工资自然就涨了。就好比一个企业, 工人工资低就招不到人, 就没法生产, 招不到人只

能提高工资，工资高了，报名的人就多了。”

观点 2. 12.26%的网友认为公务员也要养家，薪酬过低，应该加薪。

例如，网友“379866305”在谈到自己作为公务员的收入时说：“楼主 2011 年考进的公务员，湖北东南某市，房价均价 4500。每月工资 2280，扣除公积金医保后每月实发 1850，全年工资收入两万二，年底有几个月奖励性工资，大概一万三，全年所有到手收入三万五，再无其他。最近准备结婚了，顿感压力山大，到处都是用钱的，这点收入根本不够，刚去女友家认个亲都花了三万多，最近正在寻找合适时机准备辞职，在现实面前理想信念都得让步，还是去外面挣点钱现实，不为了自己，为了老婆孩子。”

观点 3. 35.10%的网友认为大多数公务员安于现状，没有抱负消极怠工，脱离群众。

例如，网友“吃吃喝喝乐呵呵 2”说：“本来已经比普通老百姓好很多了，一天干不了 5 小时，空调吹着，茶水喝着，报纸看着，不花钱的电话聊着，养老不要交钱，还嫌待遇低？这就叫贪心，将来一定就是贪官。威胁谁啊？不干了，是真的吗？那就立马走人啊！说实话就你们的那点事，社会上闯的人都比你们干得好太多了。”

观点 4. 29.65%的网友认为政府机关必须精简机构，加强公务员队伍建设，真正造福人民，在此基础上可以涨工资，不能无差别涨工资，助长不良风气。

例如，网友“labele”和“yfjxjjd”说：“先说说你认为你的老板是谁，是你的领导就让他对你满意，觉得你付出多收入少，需要加薪，你认为是纳税人，就多服务好他们，让他们觉得你不可替代，应该多付给你们薪水。”

“人民群众是主人，公务员是公仆，当然加薪要与人民满意度挂钩，否则不颠倒过来了吗！”

综上，相关详细子观点的人群分布比例如图 6.15 所示。

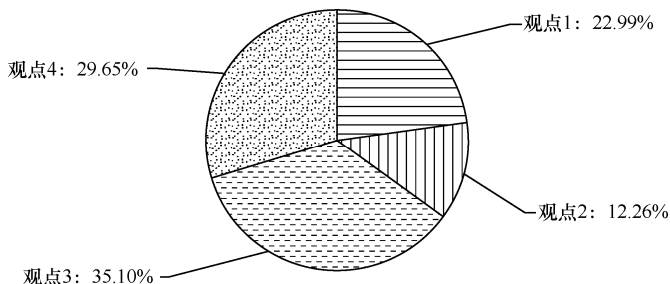


图 6.15 主题一细分子观点比例图

## 主题二：社会舆论对教师待遇问题的讨论

在本主题的进一步数据分析与研究中，发现该分类观点下还存在四个主要的细分子观点。这些细分子观点所反映的内容如下：

观点 1. 24.52%的网友认为教师待遇同工不同酬，除了存在明显地区差异，省招和县招的老师待遇也不同。

例如，网友“baifeijin”在名为“谈谈教师绩效问题”的帖子中表达了自己对于待遇不公平、不透明的质问：“① 学校是政府全额拨款单位，但是每个职工却从来不知道政府拨给学校人均人头经费是多少。职工有起码的知情权。② 所谓的绩效，领导干部的收入也从来都是不透明。③ 既然是政府的全额拨款单位，不存在盈利问题，学校的任何一个岗位都是不可或缺的，不能以绩效分成三六九等，也不应该存在一线和二线的差异，只是工种和岗位不同。④ 既然绩效了，为什么各个区的差异仍然存在。难道是各个区政府投入不一样吗。同样在上海确实收入水平不对等。重视教育的区、经济基础好的区就多。难道政府给每个区的拨款会不同吗？”

观点 2. 25.70%的网友认为如果湖北省教育厅不能处理好这次省招教师的合理诉求，必然会极大地打击年轻教师的积极性。

例如，网友“南宫羽琦 2014”在对帖子“湖北省全体省编教师致教育厅的一封信”进行评论时说：“如果湖北省教育厅不能处理好这次省招教师的合理诉求，必然会极大地打击年轻教师的积极性，导致大批年轻教师无心教学工作，一到发工资的时候就会表现的异常明显，其中新机制教师中相当一部分人会选择离开教师岗位。留下来的工作热情也会降低。而省厅后续的新教师招考也会带来诸多不利，报考者热情降低，从而导致报考人数锐减，省厅为了解决这一问题唯一的办法将会是降低招考标准，而降低标准带来的后果将会非常严重，短期来看，不利于教师的新老交替，影响教学质量，新机制将成为鸡肋，并最终成为教育厅的一大败笔。长远来说，教师队伍自身素质的降低将会对中国的基础教育产生产生很大影响。百年大计，教育为本。希望教育厅的官员们本着负责的态度处理好此事。”

观点 3. 30.25%的网友认为改变某一教师群体待遇的现状是非常难的，不一定会得到重视。

例如，网友“清香鱼”说：“改变现状还是靠自己，楼主还是虚度了很多时间，不断地学习才有进步，才能走出去，我也是在农村当了几年的教师，能理解。当年我在学校，学生上晚自习，不管值班是否我都天天去教室跟他们一起学习，在家也是把别人看电视、打麻将、聊天的时间都用来学习了，知识改



变命运，相信你能的。”

观点 4. 19.53%的网友认为教师的待遇确实比较差，给教师提高待遇能帮助教师全心投入工作。

例如，网友“喜洋洋 Z”在对帖子“湖北省全体省编教师致教育厅的一封信”进行评论是说：“中庸有云：‘凡事预则立，不预则废’，我们省招教师把我们的大好青春年华奉献给农村教育，不是我们傻，是因为这是一份值得终生付出的事业！省招教师也是省厅的新举措、新思路，我们都不希望它半途而废！我们期盼省厅能尽快调研、完善出台的各项省招生制度（包括工资增长机制等），让我们省招生在与时俱进的经济浪潮中发挥更多的光和热！”

综上，相关详细子观点的人群分布比例如图 6.16 所示。

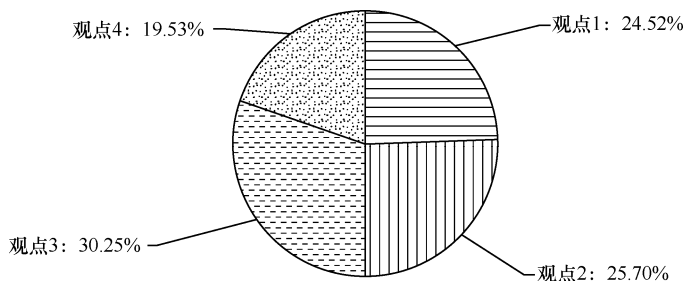


图 6.16 主题二细分子观点比例图

### 主题三：中国高物价高消费的现状和国民收入的矛盾

观点 1. 38.25%的网友认为中国进入了货币贬值、物价飞涨的时代。对中国的高物价表达了不满。

例如，网友“xuanph”说：“自认还算是社会中间层，可也倍感生活巨大压力，不知道未来会怎样。”

观点 2. 21.00%的网友认为物价飞涨，工资的涨幅远远低于物价，几乎入不敷出，在大城市生活压力巨大。

例如，网友“聪明的小龟龟”说：“房租一个劲地涨价，半年涨了 100 元。一个月的底薪不到 2000 元。房租一个月快 500 元，带上生活费，还有其他的开销根本都不够生活。说起来还是大学毕业，还不如一个中学毕业的人呢。身边的同学也都是很迷茫，不知道该找什么工作，就这样在一个心不甘情不愿的公司耗了几年，到现在工资还是屁大点，城中村也快拆了，想租个小区吧，在网上看了看，一个一室一厅都 2000 元一个月，都是吸血鬼啊。叫我们这些屌丝们

情何以堪啊。”

观点 3. 12.73%的网友认为中国的物价水平高于美国、澳大利亚，民众生活困难。

例如，网友“正红旗旗主六”以自己在美国生活的真实生活为例，说道：“美国除人工服务（如理发等很贵，美国的劳动力价格高，沾着人工的都很贵）之外，绝大部分商品的价格（绝对价格）是低于中国的，有些商品的价差之大，有点瞠目结舌。而洛杉矶的物价在美国还数比较高的，如果到俄勒冈州更便宜——购物不用缴纳 8%~10%不等的税。”

观点 4. 9.80%的网友认为美国物价并不低，以为美国物价低的人是非常无知的，中国的物价也没有美国高。

例如，网友“董长青 2013”说：“有些傻瓜一看到美国物价，比如衣服，比如小商品，比如食物，比中国便宜，立马就大呼，天啊，让我去美国受剥削去吧。或者让美国来中国剥削我吧。

什么原因造成的？这是美元全球霸权的红利。美国的核心利益，不是人权，不是领土，不是正义，是美元的世界霸权。一旦哪个国家挑战美元霸权成功，让美国失去了美元霸权，美国今后的商品价格就会突然间往上冲，到时候不光是人工，没有白送这种事情发生后，美国的商品将来也会同样贵得出奇，起码比中国这个原产地购买的时候要贵很多。”

观点 5. 18.22%的网友认为公务员工资是抵挡物价上涨的第一道防线。

例如，网友“287912227”说：“公务员工资是抵挡物价上涨的第一道防线、是稳定社会风气的最重要一块基石，公务员工资不可以涨，而且可以作为参考平衡各行业的劳动价值。我可不想，像民国末年一样，拿一麻袋的钱去买一盒火柴。”

综上，相关详细子观点的人群分布比例如图 6.17 所示。

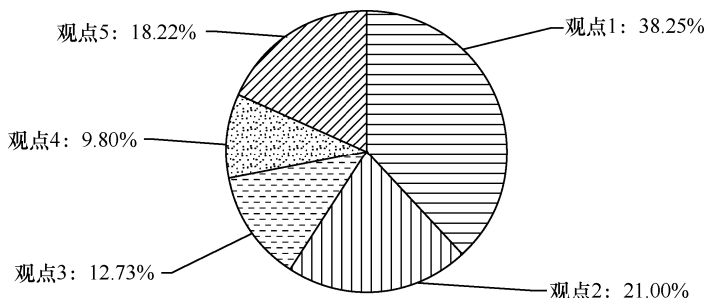


图 6.17 主题三细分子观点比例图

### 3) 观点总结

从以上的观点可以看出,在收入分配专题中,人们关注的核心问题就是个人收入以及高物价的矛盾。两大群体在收入方面的矛盾特别突出,第一类是公务员,另一类是教师。对于公务员工资问题的讨论一直是收入分配方面的热点话题。对于公务员工资是否该涨的问题,网友们形成了两大阵营,第一阵营是支持公务员涨工资,认为公务员目前的待遇偏低,第二阵营则认为公务员的工作清闲,福利好,不应该再涨工资,公务员再涨工资让其他职业的人怎么活,第二阵营的呼声更高,从这里我们可以看到公务员问题是普通民众对政府工作的态度。教师待遇的问题,每隔一段时间就会被提及,此次又引起了广大网友的热议,大部分网友认为教师的工资过低,应该提升教师的待遇。高消费高物价与普通民众低收入的情况形成了鲜明的对比,是广大网友最关心的切身利益问题,物价疯涨的速度远高于工资增长的水平,使得广大网友直呼难以接受。分配是人民休养生息的根源,合理的收入分配政策是实现社会公平的重要手段,调整贫富差距是维持社会稳定的重要条件,这些问题都值得我们给予重视。

#### 6.4.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量,专题情感分析的基础是帖子回复内容的情感趋向,涉及情感计算。

##### 1) 影响力

当一个专题确定以后,我们用专题下对应的帖子的回复数量随时间的变化来衡量专题下流量的变化。本节分别对新增回复数和总回复数按照时间进行统计,然后对统计结果进行可视化得到波线图(图 6.18,图 6.19),以便发现专题传播的发展峰值。

该专题的日流量传播图(图 6.18)中,一共达到了五次讨论热潮,这分别是由该专题下五大热帖引起的。2014年1月8日网友“忠言”发表了名为“【忠言时评】公务员退休金高于老百姓才是公平吗?”的帖子;同月23号,网友“elpasa”发表了“我也谈一谈‘公务员’的工作和收入”的长文,引发了第二次讨论热潮;此后,网友“一棵小草 909”在2014年6

月发表的帖子,“素秋 R”在 2014 年 9 月发表的帖子,以及“本末倒置 J”在 2014 年 12 月发表的帖子都引起了较大范围的热议。尤其是 2014 年 9 月,对于该专题的讨论达到了空前激烈的程度,于当月 29 日达到峰值,当天评论量达到 1311 条。

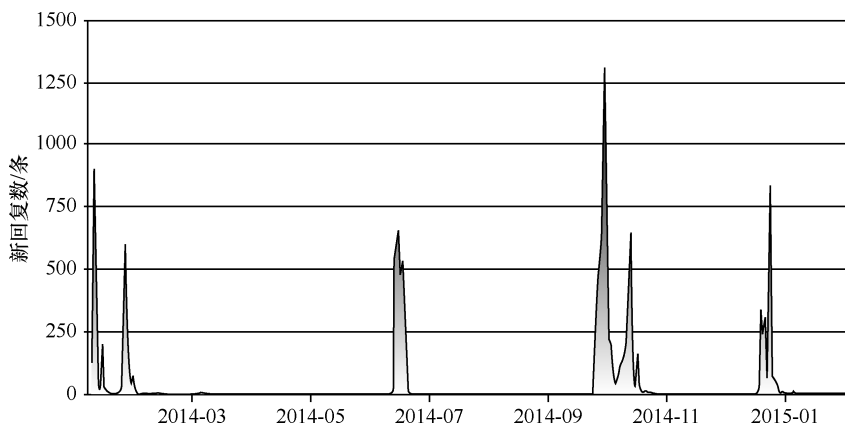


图 6.18 专题日流量时间趋势图

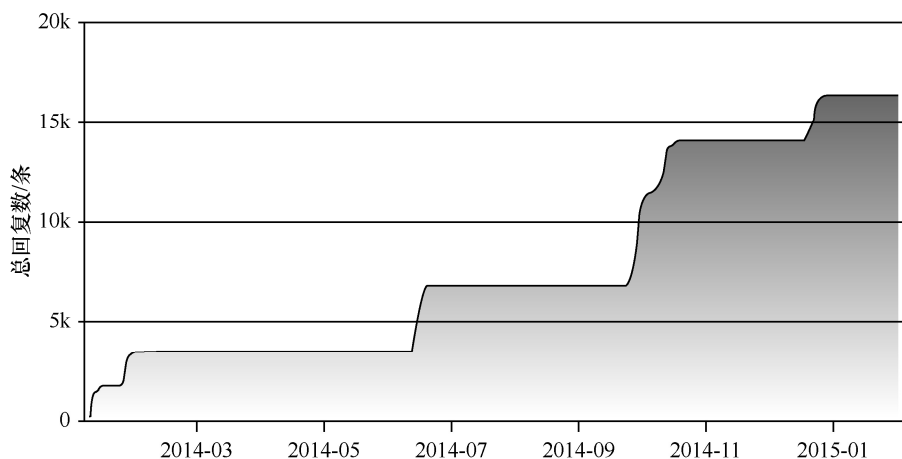


图 6.19 专题总流量时间趋势图

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库,对专题内的回复内容进行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图(图 6.20)和专题情感分类图(图 6.21)。

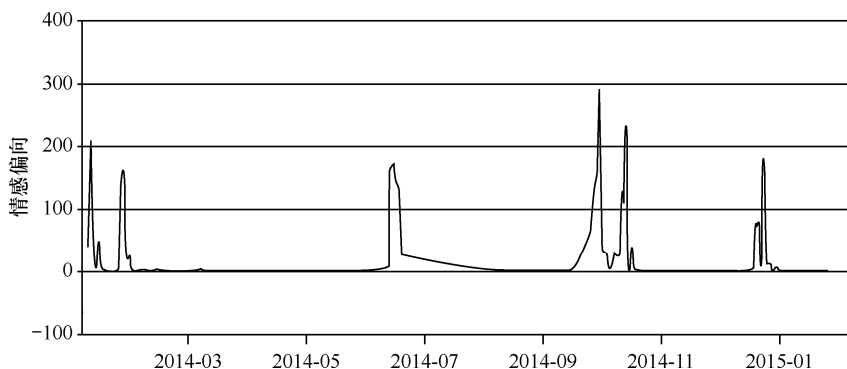


图 6.20 专题情感趋势图

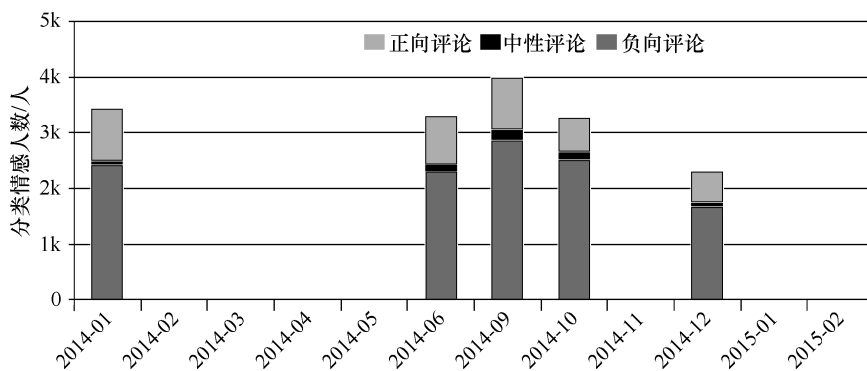


图 6.21 专题情感分类图

从图 6.20 和图 6.21 中可以看出，在 2014 年 1 月、6 月、9 月、10 月和 12 月这五个月里，随着评论数量的增加，评论的情感趋势也比较高，1 月、6 月和 10 月总评论数超过 3000 条，9 月总评论数接近 4000 条，12 月总评论数超过 2000 条，其他时间的评论数较少。负向评论的比例稳定，一直维持在总评论数的 70% 及以上。

#### 6.4.4 专题用户情感的分析

探索专题下的人物情感可以帮助我们发现该专题下的情感积极的用户，同时也可以帮助论坛找出产生消极影响的用户，剔除该用户来净化论坛环境。

本节对各个专题下的人物情感进行了统计分析，由于活跃度对情感的影响较大，所以本节构建了分别以情感值和活跃度为两个坐标轴的专题下任务的散点图 6.22。

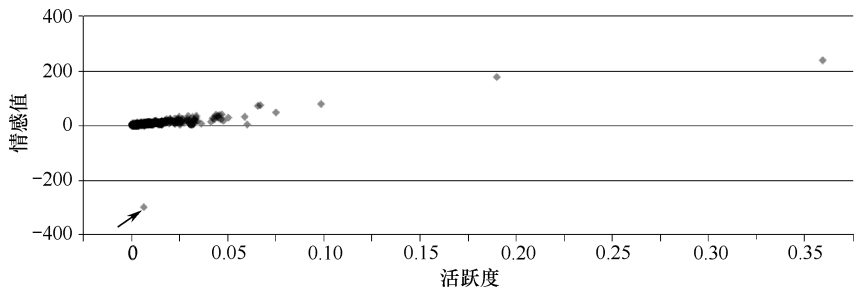


图 6.22 基于活跃度和情感值的用户散点图

圆点代表发帖人，方点代表回帖人

从图 6.22 可以看出，发帖人的情感值和活跃度都相对较低。绝大部分回帖人都集中在情感值和活跃度较低的区域。图中箭头指向的点代表的用户情感值特别低，但活跃度也很低，这类用户在网络中发表极为消极负面的言论，会在网民中产生一定的消极影响，可以采取删除言论、禁止账号等措施以消除负面影响。

为了进一步进行分析，本节将那些情感值负性较大并且活跃度相对较大的人进行了详细点的描绘，取情感值小于 20，活跃度大于 0.02 的所有用户，得图 6.23。

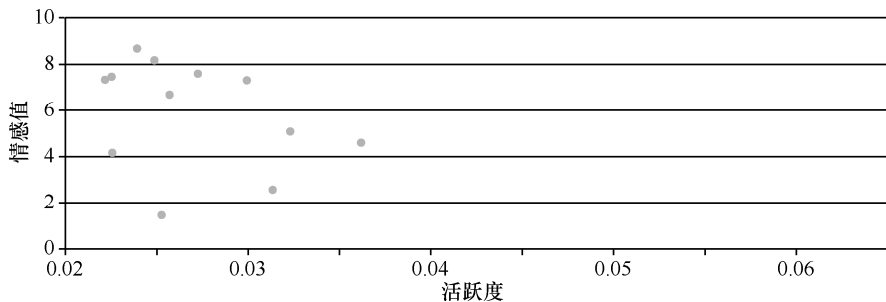


图 6.23 低情感和高活跃度散点图

图 6.23 中可以发现，在活跃度大于 0.02 的区域中并不存在情感值很负面的用户。由此可以说明，关注该事件的具有较高活跃度的用户中，不存在恶意传播消极负面信息的用户。

### 6.4.5 专题网络的构建与分析

选取收入分配专题下的相关帖子及有效评论构建专题网络。这里选取的帖

子分别为“素秋 R”“忠言”“elpasa”“本末倒置 J”“一棵小草 909”发表的有关教师和公务员收入状况的帖子。从这五篇帖子的相关评论中抽取评论关系，根据前文叙述的专题网络的构建方法构建专题网络，进行可视化得到的结果如图 6.24 所示。

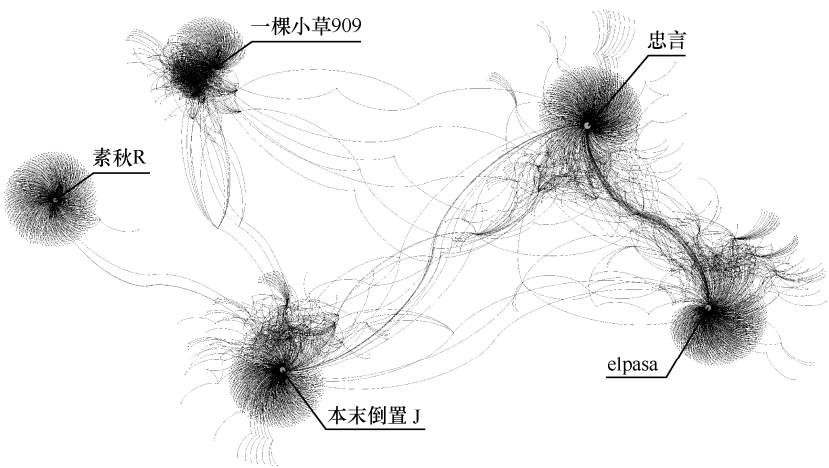


图 6.24 专题传播网络

1) 网络指标概述

这里列举了该专题传播网络的一些基本网络指标（表 6.15），使读者对该网络有一个初步的认识。该网络共由 3842 个节点和 6672 条边构成。

表 6.15 基本网络指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
1.737	0.001	18	0.709	36.064	0.096

平均度：较小，平均加权重为 4.380；图密度：该传播网络图比较稀疏；网络直径：较长，该网络的传播较远；模块化指标：划分出 5 个社区，且划分的质量较高；平均路径长度：长度较大，网络传播较广；平均聚类系数：较小。

2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算，这里取点度中心度排名前 6 的节点，列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名，得到对应的数据如表 6.16 所示。

表 6.16 中心度指标

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	值排名	绝对值	相对值	值排名	绝对值	相对值	值排名
素秋 R	7121	16.272	1	0	29.809	605	12352657	25.152	485
一棵小草 909	1707	11.898	2	264363	21.309	5	12825145	29.519	486
本末倒置 J	1429	19.500	3	0	58.007	605	2413764	38.071	1
忠言	1270	23.536	4	991878	41.322	1	5063371	33.652	14
elpasa	1159	20.672	5	609595	38.826	2	5064492	34.261	19
南宫羽琦 2014	782	0.052	6	0	0.000	605	14749441	20.099	851

从表 6.16 中可以看出“素秋 R”“一棵小草 909”等 6 个节点的点度中心度排名前 6 位。从中间中心性来看,“忠言”“elpasa”的中间中心度依次排在前两位,说明许多节点都需要经过这两个节点才能建立联系并获取信息。从接近中心性来看,“本末倒置 J”的接近中心度排名第一位,说明该节点是与网络中心最接近的节点,与其他节点进行交互比较容易。

图 6.24 的点度中心势指标为 0.24%,该指标值很小,表示该图在度数上向某点集中的趋势很不明显;图的中间中心势指标为 6.71%,该指标值不大也不小,表示该图向某些中间节点聚集的程度居中,即在整个网络中存在部分结点需要经过某些中间节点,才能得到信息。

3) 关联性与“小世界”

图 6.24 的特征途径长度为 2.854,凝聚度为 0.005。该图路径长度的具体分布如表 6.17 所示。

表 6.17 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	6607	0.005	10	69726	0.051
2	62840	0.046	11	41132	0.030
3	125420	0.091	12	20560	0.015
4	216312	0.157	13	10343	0.008
5	219139	0.159	14	2427	0.002
6	217044	0.158	15	2492	0.002
7	174710	0.127	16	501	0.000
8	125341	0.091	17	337	0.000
9	81468	0.059	18	10	0.000



图 6.24 的聚类系数为 2.940，加权的聚类系数为 0.019。节点聚类系数如表 6.18 所示。

表 6.18 节点聚类系数

用户	聚类系数	点对数
素秋 R	0.000	195000
一棵小草 909	0.007	104196
本末倒置 J	0.001	280126
忠言	0.000	408156
elpasa	0.001	314821

6.4.6 个体网络分析

按照专题领域下的整体网分析之后，我们选择了点度中心度和中间中心度最大的用户“忠言”。以该用户所发的帖子构建个体网络进行可视化结果如图 6.25 所示。

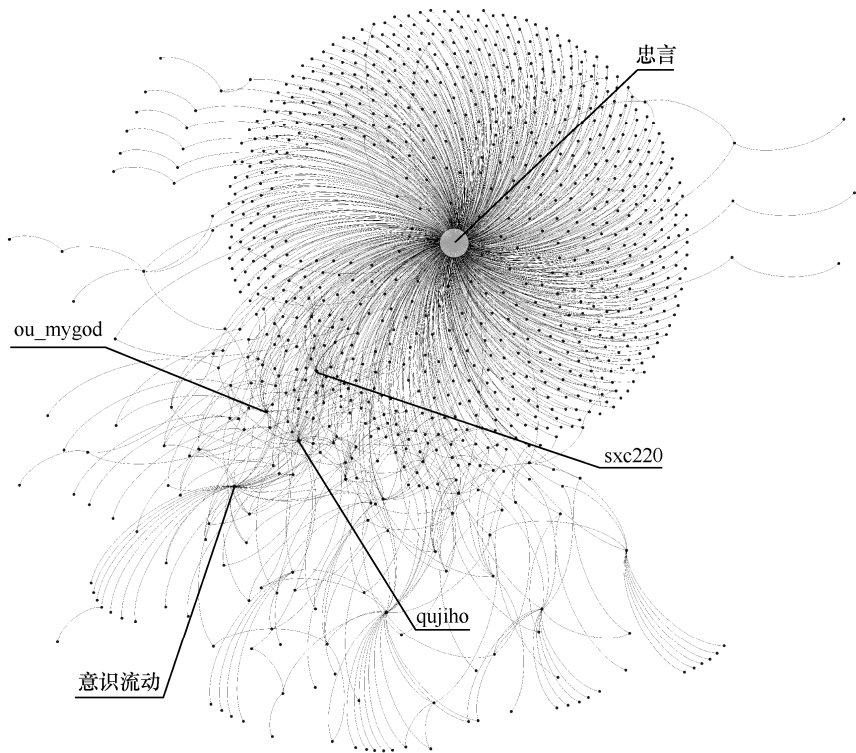


图 6.25 收入分配专题个体网络

### 1) 结构洞

这里列举节点度数排名前 5 位的用户的结构洞指标, 形成相应的数据指标表 (表 6.19)。

表 6.19 节点结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
忠言	905	891.836	0.985	0.204	0.977
qujiho	37	32.392	0.875	0.245	0.503
意识流动	34	30.630	0.901	0.213	0.682
ou_mygod	22	19.323	0.878	0.321	0.481
sxc220	18	13.158	0.731	0.427	0.474

从表 6.19 中可以看出“忠言”的节点度数、有效规模和等级度值均为最大, 其余用户则远远小于该用户, 这说明“忠言”居于网络最为核心的位置; “忠言”的效率最大, 限制度最小, 表明该节点在网络中的行动最高效, 受到网络中其他节点的限制比较小。“sxc220”对应的效率值较小, 限制度较大, 说明该节点在网络中受到其他节点的限制比较大, 行动力较弱。“意识流动”的等级度排在第二位, 位于网络中次中心的位置。

### 2) 诚实中间人

这里列举部分代表性用户的诚实中间人指标, 形成相应的数据指标表如表 6.20 所示。

表 6.20 诚实中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
忠言	28	378	119	2	1
qujiho	5	10	10	0	0
意识流动	1	0	0	0	0
ou_mygod	4	6	2	4	0
sxc220	4	6	5	1	0

HBIO 代表纯中介, HBIO1 代表弱中介, HBIO2 代表非中介。

从表 6.20 中可以看出“忠言”代表的节点共联络了 378 对关系, 其中, 该节点在 119 对关系中充当了纯中介, 在 2 对关系中充当了弱中介, 1 对关系中充当了非中介, 该节点连接的大部分关系属于纯中介, 这表明参与回复的用户大部分回复是针对主贴回复的, 也就是直接通过“忠言”相连。“qujiho”

“ou\_mygod”等节点也连接了少量节点对,在其中充当了中介。可以看出,“荣耀与卑微”发表的帖子后,大量用户参与的讨论,参与讨论的用户之间也有错综复杂的联系,因此除了中心节点“忠言”以外,其他节点充当中介的情况并不明显。

## 6.5 专题3: 社会与医疗保障

经过对民生领域的数据进行热点词和聚类分析,选取社会与医疗保障问题作为民生领域的第三个热点专题。社会保障是民生之安,关系着每个人、每个家庭的福祉。“老有所养,病有所医”既是人民最关心、最直接、最现实的利益问题,也是政府孜孜以求的目标。社会保障制度是现代国家的一项基本制度,社会保障制度是否完善已经成为社会文明进步的重要标志之一。在我国,社会保障工作直接关系到坚持党的全心全意为人民服务的宗旨,关系到维护人民群众的切身利益,关系到保证改革开放和经济建设稳定发展的大局。社会保障体系是否健全,这方面的法制是否完备,对国家的经济发展和社会稳定,会产生直接的影响。本专题中,可以看到网民对延迟退休、社保现状以及医保现状等问题进行了热烈的讨论和深入的思考。本节将对该专题下的热点事件进行进一步分析,以了解网民对相关问题的讨论、思考与见解。

### 6.5.1 专题发展简介

关于社会保障和医疗保障的民生问题一直是民众关心的热点,如退休与养老、优生优育、社保和医保等话题常常被提及。表6.21是该专题下的一些热点帖子以及相关统计指标。

表 6.21 社会与医疗保障专题热帖列表

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
(原创)70岁身体很好成延迟理由,人民代表你代表谁?	帝国良民	2014/12/26	249063	1467	717	458	23.110
嫌工资低公务员可以不辞职	剑魔独孤 2014	2014/11/21	70507	987	263	284	9.150
【自由杂谈】我爱国,国爱我吗?	自由随风 2014	2014/10/20	17168	668	209	165	73.823

续表

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
公务员加薪提议为何招致网友严厉批评?	我是蚂蚁虫	2014/3/5	75786	460	163	158	3.884
人民日报：多国公务员养老金待遇比普通人高(转载)	至中至正 2011	2014/1/6	20125	600	142	129	94.236
至中至正：工薪阶层，请不要说什么“纳税人”!	至中至正 2011	2014/1/27	23373	447	104	93	94.236
【风青杨时评】事业单位3000万人参保意味着什么?	正风青杨	2014/7/1	211684	365	131	75	62.011

如表 6.21 所示，2014 年 1 月，“至中至正 2011”在天涯论坛上连续发表了两篇关于社会保障的文章，分别为“人民日报：多国公务员养老金待遇比普通人高(转载)”和“至中至正：工薪阶层，请不要说什么‘纳税人’!”，在新年伊始便引起了广大网名的热切讨论。随后在 2014 年 3 月和 7 月，分别又引起了一轮新的关于公务员社会保障乃至事业单位人员参加社保的热议。2014 年 12 月，“帝国良民”发表了“70 岁身体很好成延退理由，人民代表你代表谁？”一帖，一跃成为新的热帖，该帖子在很短的时间内引起网友的大量回复，就工人延迟退休问题进行了激烈的讨论。

在帖子的专题传播过程中，大量的网民针对帖子所表达的观点进行了回复，其中也产生了一些颇具影响力的新观点，并引发了网名的关注的讨论。

6.5.2 观点分类与观点列表

通过上述的网络数据信息，针对网民在回复过程中的观点挖掘与倾向进行语义的处理，形成了本话题内的核心观点。

1) 观点分类

根据观点涉及的特征词汇和领域以及网民发表的各个观点内容本身之间存在的相关性与差异性，可以将所有评论分为以下三个主题，各个主题的观点比例图如图 6.26 所示。

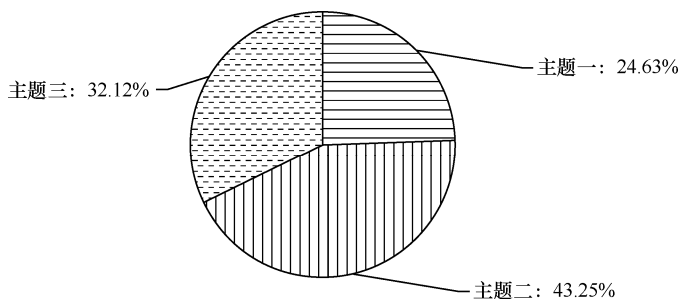


图 6.26 “社会与医疗保障”观点分类比例图

### 主题一：关于延迟退休问题的热议

此分类观点包含三个关键特征词：社会老龄化、延迟退休、提前退休。对应的观点包括：

观点 1. 延迟退休是自欺欺人、饮鸩止渴的办法，应该鼓励提前退休，加速新陈代谢。

观点 2. 延迟退休是应对社会老龄化的必然性结果，有退休已经不错了。

观点 3. 关于退休年龄的问题不能一刀切，不能听所谓专家、代表的退休怪论。提出延迟退休的人缺乏基本的人性和公德，无视中国社会的现实。

### 主题二：社会保障问题的现状

此分类观点包含三个关键特征词：社保、利益、财政。对应的观点包括：

观点 4. 事业单位统一参加社保，是社会保障体制的进步。

观点 5. 政府在社会保障中能且只能承担有限责任，财政投入社会保障应该量力而行，不能无限制的填这个无底洞。

观点 6. 社会保障的改革应该实现公平原则，不能总是老人老办法，新人新办法，保护部分既得利益者。

### 主题三：医疗保障问题的现状

此分类观点包含三个关键特征词：医保、公平、特权。对应的观点包括：

观点 7. 医保政策存在很大问题，全国的专科医院不联网，异地看病难报销。

观点 8. 现在的医保卡被当成购物卡使用，医保套现的情形屡禁不止，医保沦为部分人员的利益工具。

观点 9. 中国医保成鸡肋，得了大病医保不管或报销比例极少，小病大医的情况也屡见不鲜。

观点 10. 适当添加商业保险补充是很有必要的，解决看病问题不只靠国家，也要靠自己。

## 2) 观点分类详细分析

为了更好地从宏观向微观的深入分析,本节对每个分类观点内的网络用户分布以及网络民意进行了进一步的分析,希望通过网络用户真实的反馈信息,来反映他们对这一些问题的思考与建议。

### 主题一: 关于延迟退休问题的热议

在进一步的数据分析与研究中,对网民在这个主题内的深入观点进行梳理与分析,发现存在3个细分的子观点。这些细分的子观点所反映的内容如下:

观点 1. 32.12%的网友认为延迟退休是自欺欺人、饮鸩止渴的办法,应该鼓励提前退休,加速新陈代谢。

例如,网友“任随忽悠”说:“可以肯定地讲,在就业形势日益严峻、大量劳动力严重过剩的情况下,提出延期退休是不明智的,不可取的。明智聪明的做法是,鼓励适当提前退休,提前退休期间领取50%或者60%的退休金,达到法定退休年龄时,领取全额退休金;凡是有提前退休的家庭,如有适龄就业子女而没有就业,政府优先安排工作,特别是大学生子女更要优先安排。至少,我们应当把是否提前退休交给劳动者自愿选择,给劳动者一个选择的机会,而不要那样蛮横,那样僵硬。”

政府手中的钱即财政收入,本质上是劳动者委托政府管理、支出的钱。这笔钱怎么花,每年拿出多少钱养老人,从根本上讲,应该是全社会的劳动者说了算。总而言之,笔者认为,我国劳动力将长期处于过剩状态,延期退休不可取。我国养老问题、养老金问题没有人们想象的那样严重。只要政策对头,我国未来60岁老人将领取更多的养老金,生活将更幸福。”

观点 2. 15.54%的网友认为延迟退休是应对社会老龄化的必然性结果,有退休已经不错了。

例如,网友“青山依旧毕竟东流”直接表达了自己无奈:“延迟退休,必然性的结果,意料中的事,有退休已不错了”。

观点 3. 52.34%的网友认为关于退休年龄的问题不能一刀切,不能听所谓专家、代表的退休怪论。提出延迟退休的人缺乏基本的人性和公德,无视中国社会的现实。该观点针对的是人大代表苏晓云提出70岁身体很好、不需要退休,以及清华教授杨燕绥提出65岁领养老金、没领养老金时男当园丁女洗衣服的言论,引起了网友的普遍反感。

例如,网友“tlsfz”和“caoyan123”表达了自己的质疑和思考:“饱食终日者始终不知基层劳动者之辛劳,让他(苏晓云)也在65~70岁时干点体力劳动来填饱肚子,看他能撑住?”

“政府让一些脱离中国社会现实的迂腐蠢货作为所谓的智囊，实在令人担心。但毕竟最终政策是要政府定的，如果真的采纳了如此缺德的清华方案，那才真是国民的‘杯具’；‘叫兽’可以闭门造车，国策可不能当儿戏，毕竟每个人要老的。”

综上，在本主题中，具有相应子观点的评论分布比例如图 6.27 所示。

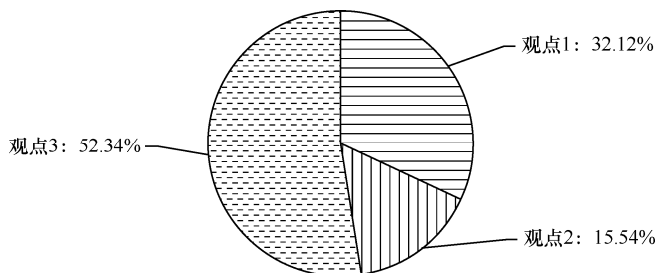


图 6.27 主题一细分子观点比例图

## 主题二：社会保障问题的现状

在进一步的数据分析与研究中，对网民在这个主题内的深入观点进行梳理与分析，其中在本主题下，存在 3 个细分的子观点，这些细分的子观点所反映的内容如下：

观点 1. 30.23%的网友认为事业单位统一参加社保，是社会保障体制的进步。

例如，网友“江津集邮社”说：“长期以来，事业单位在养老方面‘独树一帜’，拥有一些特权，不参加社保，由公共财政买单，造成的结果是，公共财政的负担加重，事业单位退休人员比企业单位退休人员的待遇高不少，这长期被社会所诟病，也是实实在在的社会不公。今后，事业单位工作人员也将参加社保的规定，是剔除特权、维护社会公平的表现。”

观点 2. 44.11%的网友认为政府在社会保障中能且只能承担有限责任，财政投入社会保障应该量力而行，不能无限制的填这个无底洞。

例如，网友“至中至正 2011”说：“政府不能包打一切，不能包办劳动者的养老问题，甚至不能承担养老的主要责任。养老问题主要是劳动者自己的责任，政府主要负责职工缴纳的养老金的保值和安全。政府对于社会是管理机构，不是生产单位，本身并不会凭空创造更多的财富。所谓取之于民用之于民，社会保障的资金来源，归根结底还是来源于社会，不是天上掉馅饼，也不是政府凭空创造的。所以，财政投入社会保障的资金，应该量力而行。要避免打肿脸充胖子，财政本身已经没钱了，还一再上调；现在没钱，挪用年轻一代缴纳的养老金来补窟窿，更有甚者，政府发行国债填补社保的窟窿。要是这样一来的话，估计社保基金体系迟早会崩溃。”

观点 3. 25.66%的网友认为社会保障的改革应该实现公平原则，不能总是老人老办法，新人新办法，保住部分既得利益者。

例如，网友“扩展 KZ”说：“事业单位改革实则是想甩掉一些包袱，保住部分既得利益者。从表面上似乎消除特权打造公平，实际上是抛弃的是在编的一线职工，也就是将一线职工利用潜规则的方式推进企业的行列，而处于二线的机关管理人员仍旧维持不动，也就是维持部分人的所谓理所应当的特权待遇，所以说换汤不换药，丢卒是改革的成果，保车才是目的。”

综上，在本主题中，具有相应子观点的评论分布比例如图 6.28 所示。

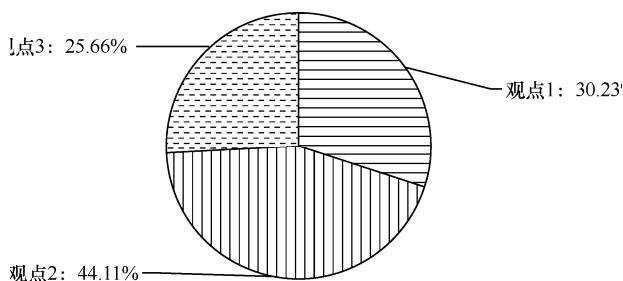


图 6.28 主题二细分子观点比例图

### 主题三：医疗保障问题的现状

在进一步的数据分析与研究中，对网民在这个主题内的深入观点进行梳理与分析，发现存在 4 个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 24.29%的网友认为医保政策存在很大问题，全国的专科医院不联网，异地看病难报销。

观点 2. 31.51%的网友认为现在的医保卡被当成购物卡使用，医保套现的情形屡禁不止，医保沦为部分人员的利益工具。

例如，网友“2xiaomi”在帖子“谈谈医保下的百态众生，闪烁着丑陋灵魂和贪得无厌！”中说：医保本身的出发点是好的，本身的宗旨也值得赞扬的。但在全民医保这块蛋糕上面，很多人做起了文章，甚至很多人骗取医保。这些丑陋的行为迫使医保改革需要一个监督和制度来保障。不然的话，医保成为一种伤害民心的一个利器。毕竟现在人民生活水平提高了，对健康意识加强了。需要的医疗服务水平也相应提高了。在这种情况下，医疗事业的发展跟不上人民群众的需求，矛盾巨大。这些现实问题制约着社会人民生活水平的进一步提高。



观点 3. 34.22%的网友认为中国医保成鸡肋，得了大病医保不管或报销比例极少，小病大医的情况也屡见不鲜。

例如，网友“三字经是个好东西”在“中国医保之惑：大病治不起，小病不屑治”一帖中说：“医生缺德，国家缺钱，百姓医闹，真一个难字了得。”

观点 4. 9.98%的网友认为适当添加商业保险补充是很有必要的，解决看病问题不能只靠国家、也要靠自己。

例如，网友“香港安盛 101”说：“社会保险和商业保险，同出一门，互为补充，谁也替代不了谁。社保是我们生活最最基本的保障，是维持基本生活的保证。而商保强大、独特的保障功能和资产保全功能是所有其他任何金融工具不可替代的。所以，要保障我们的整体生活质量，任何时候都能从容面对生活中的起伏与风雨，仅有社保是不够的，必须再加上商业保险，才能保生活的完美、富足！”

综上，在本主题中，具有相应子观点的评论分布比例如图 6.29 所示。

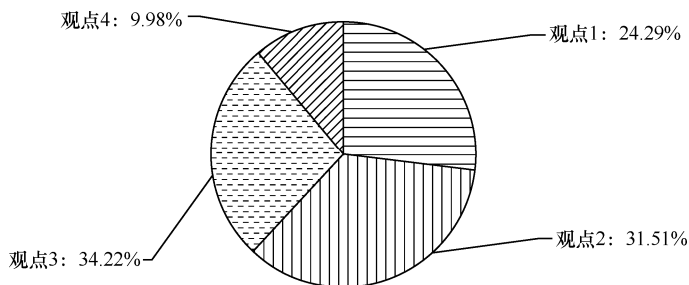


图 6.29 主题二细分子观点比例图

### 3) 观点总结

从网友对于“社会与医疗保障”问题的讨论中可以看出，网友对社保和医保的现状进行了热烈的讨论，表达了各自对现有政策的见解和意见，普遍认为现有政策虽然从一定程度上解决了社会问题，但是也存在诸多弊端，需要改进。另外，关于工人退休年龄的问题引发了一场全民热议，这源于一位人大代表提出的延迟退休的建议，引起了广大民众的不满。“社保”是人民生存和发展的依托，就是要把老百姓都装进“保险箱”，完善和健全养老、失业、医疗等社会保障机制，落实城镇居民最低生活保障；探索建立农村养老、医疗保险和最低生活保障制度；大力加强对特殊困难群众的救助，确保弱势群体的生活底线，使人民群众老有所养，病有所医，居有其屋，衣食无忧。

### 6.5.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

#### 1) 影响力

当一个专题确定以后，我们用专题下对应的帖子的回复数量随时间的变化来衡量专题下流量的变化。本节分别对新增回复数和总回复数按照时间进行统计，然后对统计结果进行可视化得到波线图（图 6.30，图 6.31），以便于发现专题传播的发展峰值。

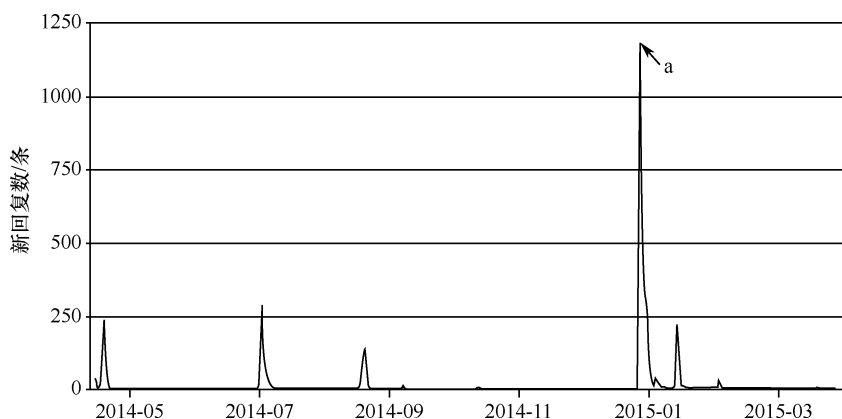


图 6.30 专题日流量时间趋势图

点 a：2014 年 12 月 26 日

从图 6.30 可以看出，每天对应的新增回复数的变化。在点 a 之前，每天对应的新增回复量非常少，说明这一段时间该事件提出来之后并没有引发大的影响，只在小范围内引起了三次集中讨论；2014 年 12 月 26 日，用户“帝国良民”发表了一篇名为“70 岁身体很好成延退理由，人民代表你代表谁？”的帖子，在当天便引起了 1178 条回复，达到了峰值，随后几天，该帖子的关注慢慢减退；在 2015 年 1 月 12 日，又达到了一次小高潮。在很短的时间内引起网友的大量回复，就工人延迟退休问题进行了激烈的讨论。从专题总流量趋势图（图 6.31）可以明显看到几次回复数骤增的节点，发现讨论热点。

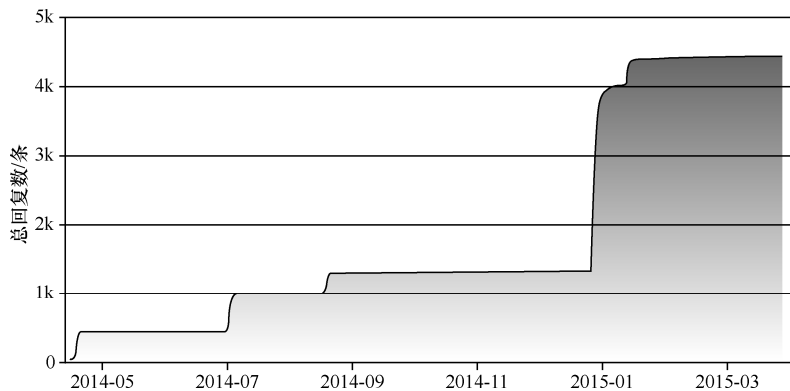


图 6.31 专题总流量时间趋势图

2) 传播情感分析

本节利用 **HowNet** 建立一个有效的中文情感词库，对专题内的回复内容进行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图（图 6.32）和专题情感分类图（图 6.33）。

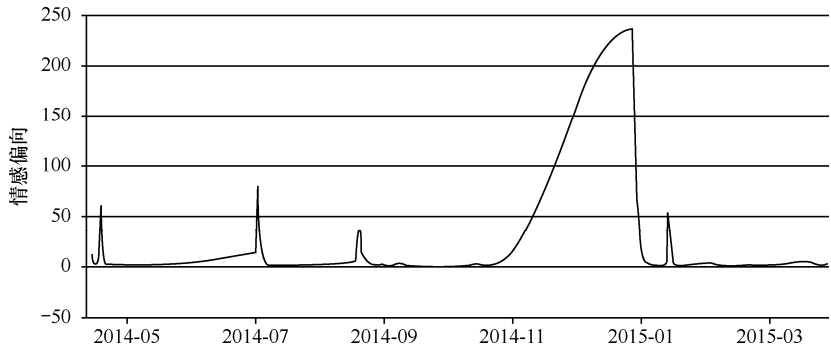


图 6.32 专题情感趋势图

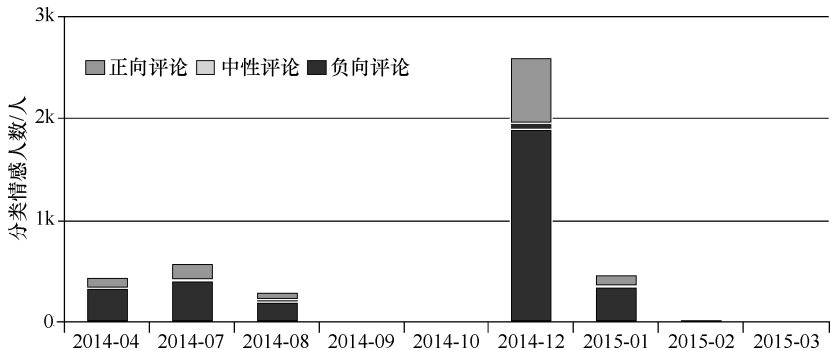


图 6.33 专题情感分类图

从图 6.32 可以看出事件发展的拐点。从大体上来看，专题情感趋势图（图 6.32）和日流量统计图（图 6.30）的趋势基本一致。这是由于进行情感统计的方式是累加，而且一般正向评论的值都比较大，负向评论的值都比较小，累加计算得到的值以正数居多，评论数越多，统计规律越明显。但是从专题情感分类图（图 6.33）中看到，负向情感的评论数量远大于正向评论，以 2014 年 12 月为例，正向评论、中性评论和负向评论的数量分别为 637、66 和 1895，所占比例分别为 25%、3%和 73%。

6.5.4 专题用户情感分析

探索专题下的人物情感可以帮助我们发现该专题下的情感积极的用户，同时也可以帮助论坛找出产生消极影响的网民，剔除该用户来净化论坛环境。

本节对各个专题下的人物情感进行了统计分析，由于活跃度对情感的影响较大，所以本节构建了分别以情感值和活跃度为两个坐标轴的专题下任务的散点图（图 6.34）。

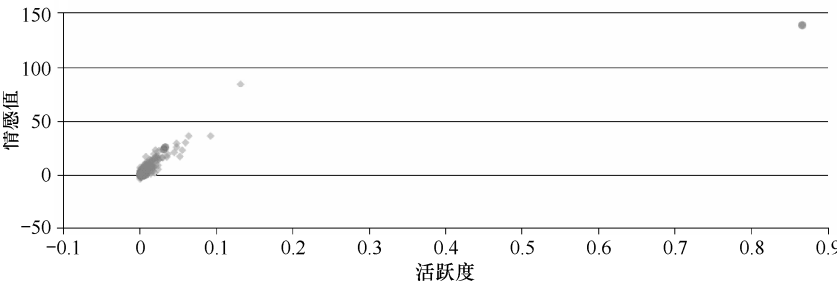


图 6.34 基于活跃度和情感值的用户散点图  
圆点代表发帖人，方点代表回帖人

从图 6.34 可以看出，本专题中，其中一个发帖人的情感值和活跃度远远大于其他用户，该用户可以视为本专题中的意见领袖。其余的发帖人和回帖人都集中在情感值和活跃度较低的区域，是论坛中最大最普遍的群体。

为了进一步进行分析，本节将那些情感值负性较大并且活跃度相对较大的人进行了详细点的描绘，取情感值小于 20，活跃度大于 0.02 的所有用户，得图 6.35。

图 6.35 中可以发现，活跃度大于 0.02 且情感值小于 20 的区域中的用户数量比较少，且情感倾向都比较正向。由此可以说明，关注该事件的用户中，具有较高活跃度的用户情感倾向都比较正面。

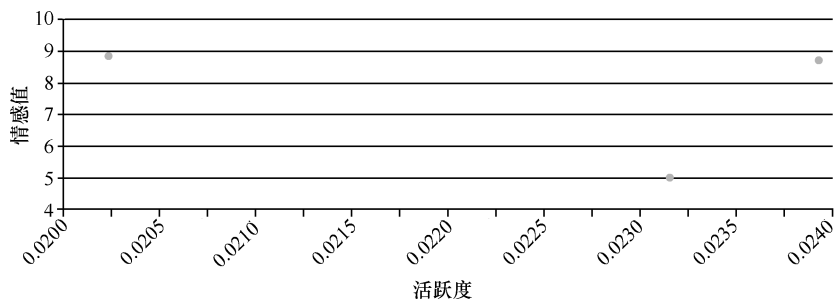


图 6.35 低情感和高活跃度散点图

6.5.5 专题网络的构建与分析

选取社会与医疗专题下的相关帖子及有效评论构建专题网络。这里选取的帖子分别为“帝国良民”“正风青杨”“别了还来”“钱塘超人在漂泊”“云卷云舒 321a”发表的有关社会保障和医疗保障现状的帖子。从这五篇帖子的相关评论中抽取评论关系，根据前文叙述的构建专题网络的构建方法构建专题网络，进行可视化得到的结果如图 6.36 所示。

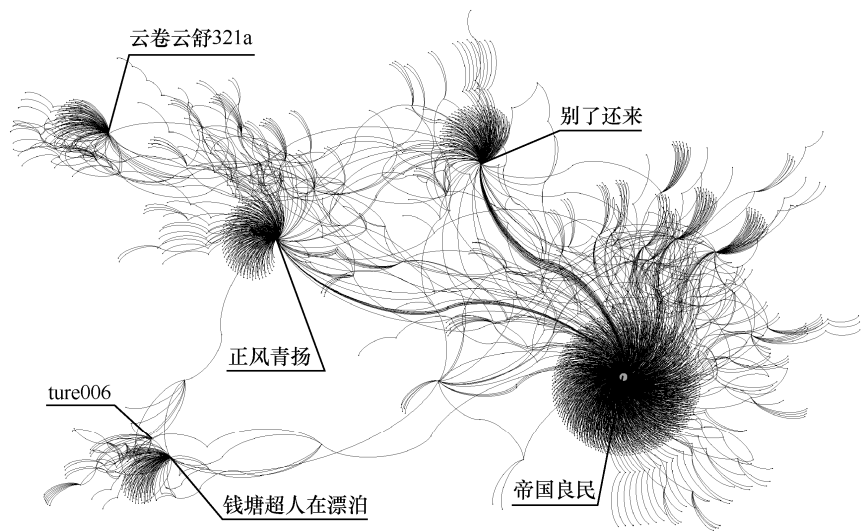


图 6.36 专题传播网络

1) 网络指标概述

这里列举了该专题传播网络的一些基本网络指标（表 6.22），使读者对该网络有一个初步的认识。该网络共由 2839 个节点和 3499 条边构成。

表 6.22 基本网络指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
1.232	0.001	22	0.611	5.423	0.046

平均度：较小，平均加权重为 1.591；图密度：该传播网络图比较稀疏；网络直径：较长，该网络的传播特性比较明显；模块化指标：划分出 10 个社区，且划分的质量较高；平均路径长度：居中，网络传播较广；平均聚类系数：很小。

2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算，这里取点度中心度排名前 6 的节点，列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名，得到对应的数据如表 6.23 所示。

表 6.23 中心度指标

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	值排名	绝对值	相对值	值排名	绝对值	相对值	值排名
帝国良民	2079	57.188	1	—	85.978	—	2582956	25.847	1
正风青杨	357	11.135	2	0	24.923	302	5439479	20.382	2
别了还来	321	8.809	3	18264	16.631	3	6406483	20.194	4
钱塘超人在漂泊	212	3.559	4	5322	8.216	8	7628604	18.499	54
云卷云舒 321a	190	3.629	5	24673	7.300	1	6407728	17.160	8
ture006	138	0.564	6	1179	0.367	28	7628731	15.993	55

—表示该指标不存在。

从表 6.23 中可以看出“帝国良民”、“别了还来”等 6 个节点的点度中心度排名前 6 位。从中间中心性来看，“云卷云舒 321a”的中间中心度远远大于其他节点，排在第一位，说明许多节点都需要经过该节点才能建立联系并获取信息。从接近中心性来看，“帝国良民”和“正风青杨”的接近中心度排名前两位，说明这两个节点与其他节点进行交互比较容易。

图 6.36 的点度中心势指标为 0.31%，该指标值很小，表示该图在度数上向某点集中的趋势很不明显；该图的中间中心势指标为 0.92%，该指标值也比较小，表示该图向某些中间节点聚集的程度比较小，即在整个网络中大部分的结点并不需要经过某些中间节点，就能得到信息。

3) 关联性与“小世界”

图 6.36 的特征途径长度为 5.423，凝聚度为 0.002。该图路径长度的具体分布如表 6.24 所示。

表 6.24 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	3479	0.080	12	398	0.009
2	5352	0.124	13	549	0.013
3	5693	0.132	14	677	0.016
4	4391	0.102	15	435	0.010
5	3379	0.078	16	228	0.005
6	7936	0.183	17	260	0.006
7	5419	0.125	18	161	0.004
8	2806	0.065	19	92	0.002
9	914	0.021	20	47	0.001
10	576	0.013	21	17	0.000
11	441	0.010	22	5	0.000

图 6.36 的聚类系数为 1.005，加权的聚类系数为 0.01。节点聚类系数如 6.25 表所示。

表 6.25 节点聚类系数

用户	聚类系数	点对数
帝国良民	0.000	1316253
正风青杨	0.000	49770
别了还来	0.001	31125
钱塘超人在漂泊	0.012	5050
云卷云舒 321a	0.006	5253

6.5.6 个体网络分析

经过之前对整体网的分析，本节从中选取那些节点度数相对较大、传播性比较好的帖子，做了该帖子下的参与的所有的用户的个体网络。本节选取的是“帝国良民”发表的一篇帖子，对它进行指标的分析，并且描述传播路线。

1) 结构洞

这里列举节点度数排名前 5 位的用户节点的结构洞指标，形成相应的数据指标表（表 6.26）。

表 6.26 节点结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
帝国良民	1623	1622.668	1.000	0.002	0.398
星星的眼睛 A	56	55.262	0.987	0.049	0.380
七喜如意 2014	48	47.506	0.990	0.052	0.342
a500miles	2	1.024	0.512	0.978	0.976
哥们姐妹玩家家	13	4.783	0.368	0.801	0.741

从表 6.26 中可以看出“帝国良民”的节点度数、有效规模和效率均为最大，限制度最小，这说明“帝国良民”在网络中和绝大多数其他节点都有联系，行动最高效，受到网络中其他节点的限制比较小。“a500miles”对应的等级度最大，效率值较小，限制度较大，说明该节点在网络中位于核心位置，且受到其他节点的限制比较大，自由度很弱。

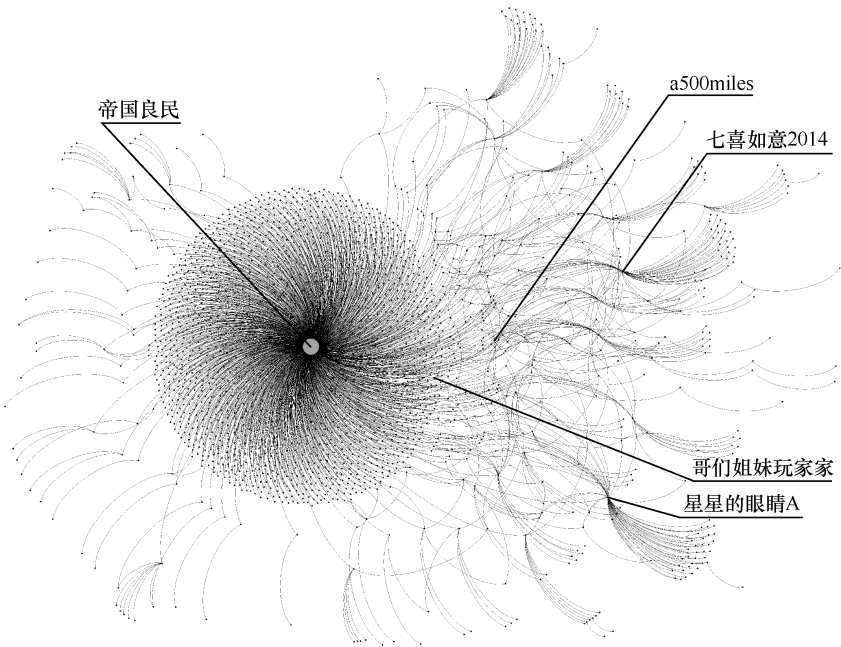


图 6.37 社会与医疗保障专题个体网络

2) 诚实中间人

这里列出部分代表性用户的诚实中间人指标，形成相应的数据指标表如表 6.27 所示。



表 6.27 诚实中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
幺鸡二条不打要遭	2	1	1	0	0
迎面吹来西北风	2	1	1	0	0

HBIO 代表纯中介，HBIO1 代表弱中介，HBIO2 代表非中介。

从表 6.27 中可以看出，除了“幺鸡二条不打要遭”和“迎面吹来西北风”分别联系了一对节点，充当了一次纯中介外，其余用户该项指标全为 0，这说明该话题在传播过程中，几乎很少出现互动的情形，而是信息沿着单一方向进行扩散，发帖作者与评论者也没有发生任何交流。

## 6.6 本章小结

本章主要针对天涯论坛的民生领域的舆情进行了分析，选取了网民关注最广泛的教育与就业、收入分配和社会与医疗保障三个专题，分别从内容和结构上进行了深入的研究和分析。在内容分析方面，通过观点挖掘、影响力分析和情感分析等，我们针对相关的网络舆情进行了分析，包括关注该领域的网民对民生领域主要问题的一些观点以及情感态度。此外，通过构建民生领域整体网、高频词时间演化网、专题传播网络等社会网络，应用社会网络分析法，直观地发现和展示了一些与内容分析相符的结论。并通过建立基于专题的传播社会网络，发现在该专题发展形成过程中起到舆论引导作用的主要用户信息。

党的十八大报告关于保障和改善民生的重要论述指出：“提高人民物质文化生活水平，是改革开放和社会主义现代化建设的根本目的。”从这个意义上说，无论是全面建成小康社会、实现中华民族伟大复兴宏伟目标，还是经济、政治、文化、社会、生态“五位一体”总体布局，实际上都贯穿着切实关注民生、重视民生、保障民生、改善民生这条主线，都以努力让人民过上更好的生活为目标和归宿。

由于民生领域涉及百姓的教育、就业、收入和保障等与老百姓个人切身利益紧密相关的一系列问题，因而一直受到广泛的关注。教育是强国富民的基础，就业是人民生存和生活的根本，教育资源分配不均、中国教育的现状、大学生就业和创业问题是教育与就业专题中的热点话题。合理的分配是要实现改革发展成果让人民共享，公务员薪资问题、教师待遇问题以及中国高物价与低工资的矛盾涉及社会的分配体系的公平性。工人退休问题以及当前的

社会和医保现状是社会和医疗保障专题中的热点话题。本章以上述各个专题下的热点话题为切入点进行内容分析，可以管中窥豹，了解网民对民生领域的舆情概况。

民生连着民心，民心连着党心。只有着力改善民生，把人民群众最关心、最直接、最现实的利益实现好、维护好、发展好，才能使党同人民息息相通。“忧民之忧者，民亦忧其忧；乐民之乐者，民亦乐其乐。”只要始终关注民生、不断改善民生，就可以凝聚民心、集中民智、发挥民力，才能顺利实现“两个一百年”奋斗目标和中华民族伟大复兴的中国梦！

## 第 7 章 生态领域的相关舆情分析

20 世纪 50 年代以来,随着科技的突飞猛进和经济开发的扩大化、区域化、集中化,人类利用和改造自然的能力和规模空前增强,排放的工业废弃物也迅速增加,甚至酿成了许多震惊世界的环境公害事件,形成了环境问题的第一次高潮。但是,这些都是在发达国家中由于工业污染造成的区域性环境问题,人们关心的一般是对人体健康的影响,生态破坏问题等尚未引起重视。

70 年代以后,环境问题愈演愈烈,一方面由于人口过快增长和不可持续的消费方式,使人们掠夺式地开发自然资源,造成资源枯竭和生态破坏;另一方面又向自然环境排放大量工业和生活废弃物,破坏了自然界物质系统的正常循环方式,这些物质的扩散、迁移、积累甚至可能影响到全球。温室效应引起的全球气候变暖、臭氧层损耗、酸雨区的扩展、自然资源的耗竭、大面积生态破坏等引发了环境问题的第二次高潮,这次高潮的影响范围在广度和深度上都远远超过了五六十年代的那一次。目前的环境问题已不仅仅是对人体健康的影响,而是已极大地削弱了整个地球生命支撑系统的支持能力。人类如不采取有效措施控制环境污染和生态破坏不断加剧的趋势,人类社会的生存基础将被瓦解和动摇,环境问题的实质已经成为生存问题。

针对这些问题,我国相继出台了一系列相关的法律和法规,并在 2012 年 11 月党的十八大中提出了“大力推进生态文明建设”的战略决策,希望通过建立系统完整的生态文明制度体系,实行最严格的源头保护制度、损害赔偿制度、责任追究制度,完善环境治理和生态修复制度,用制度和法律来保护生态环境。特别是 2013 年 11 月召开的十八届三中全会上,进一步明确了生态改革的目标,其中包括:制定实施生态文明建设目标体系、推进生态保护红线划定工作、深化环评审批制度改革、建立健全生态补偿机制、完善排污许可制度和企事业单位污染排放总量控制制度、完善环境政策法规等。这为未来生态改革的具体方针政策制定指明了工作方向。方针政策的实施需要广大民众的积极参与,普通民众对环保问题上有什么样的建设思考,以及他们对环境保护与生态文明建设所持有的态度和倾向都会影响到政策的实施结果,因此获取生态领域网民的观点并进行总结分析就会为我们客观了解民众对这些政策的态度提供了保障。

本章从天涯论坛选取了生态领域相关的数据，特别是针对生态领域中的热点专题进行深入的总结和分析，从而了解网民在生态方面建设的舆情。

7.1 生态领域热点专题和热点词分析

本节选取生态领域的相关帖子进行分析，通过对词频进行统计得到该领域的热点词及演化情况，通过对原帖进行文本聚类得到热点事件，从而确定该领域的热点专题。

7.1.1 生态领域热点词分析

热点词出现的频率会随着热点事件的发生而快速增长，从热点词的演化过程可以看出不同时间热点事件的主题变化情况。对 2014 年全年的数据按季度进行词频统计和排名，结果如表 7.1 所示。

表 7.1 高频词汇统计

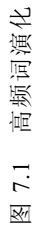
2014.01~2014.03		2014.04~2014.06		2014.07~2014.09		2014.10~2014.12	
词语	词频	词语	词频	词语	词频	词语	词频
污染	1910	污染	2384	污染	2527	污染	2903
环保	1048	环保	1519	能源	1567	环保	1949
能源	923	能源	1400	环保	1557	能源	1173
石油	642	石油	1121	石油	1263	环境污染	921
环境污染	469	天然气	632	天然气	858	石油	801
空气污染	341	节能	536	环境污染	518	空气质量	529
环境保护	329	环境污染	508	污水	477	牧业	507
天然气	321	水源	488	发电	469	大气污染	487
燃料	317	生态环境	452	节能	446	节能	459
生态环境	298	发电	442	水源	442	天然气	433
节能	297	污水	416	环境保护	394	生态环境	418
污水	290	环境保护	405	生态环境	345	减排	413
发电	290	地质	356	污染物	321	污染物	410
污染物	274	燃料	315	水污染	302	环境保护	405
空气质量	263	新能源	304	燃料	292	发电	366
水源	258	污染物	301	空气污染	283	污水	363
大气污染	248	饮用水	289	饮用水	282	空气污染	300
地质	244	空气污染	266	地质	265	水源	276

续表

2014.01~2014.03		2014.04~2014.06		2014.07~2014.09		2014.10~2014.12	
词语	词频	词语	词频	词语	词频	词语	词频
尾气	197	原油	241	太阳能	245	燃料	270
减排	160	水资源	229	原油	232	原油	257
新能源	157	大气污染	220	污水处理	223	新能源	244
气温	150	植被	202	空气质量	223	尾气	219
林业	148	太阳能	194	新能源	219	林业	198
饮用水	148	林业	193	水资源	218	锅炉	186
太阳能	142	空气质量	192	污染源	188	地质	178
水资源	140	气温	191	大气污染	180	污水处理	173
电池	138	采矿	191	保护环境	179	水资源	167
反应堆	138	污水处理	190	气温	179	太阳能	159
保护环境	131	减排	180	林业	178	节能减排	151
环境问题	129	电池	173	采矿	162	水污染	142
植被	129	干旱	172	流域	152	饮用水	142
污水处理	121	水污染	171	环境问题	148	电池	138
污染源	119	流域	151	自然环境	148	颗粒物	137
水污染	118	保护环境	138	电池	145	植被	137
自然环境	118	垃圾处理	137	减排	142	气温	128
流域	117	污染源	133	尾气	135	污染源	126
原油	114	造林	130	多样性	134	二氧化硫	124
废气	112	节能减排	129	废气	132	环境质量	120
颗粒物	110	环境问题	128	森林公园	131	保护环境	115
放射性	107	共生	125	锅炉	127	造林	110

为了更好地观察高频词的演化过程,我们用作图工具 NetDraw 对高频词和时间的对应关系进行可视化,得到了高频词演化图(图 7.1)。

从热点词和词频来看,“大气污染”“水污染”“节能”“减排”等词一直稳定的出现在每个季度之中,这些稳定存在的词既反映了 2014 年生态领域网民一直比较关注的问题,如“空气污染”“水污染”等;也反映了网民对于这些问题一直持有的态度及措施,如“节能”“减排”“新能源”等。除此之外,还有些一直在变化的词,有的只在一个季度出现,有的在两个或三个季度同时出现。“反应堆”“放射性”这些热点词只在第一季度出现,说明了福岛核电站放射性



液体泄漏事件对于生态领域的影响比较大;“森林公园”“多样性”等热点词只在第二季度出现应该和 5 月 22 日的国际生物多样性日有关;在第三季度出现的“干旱”热点词说明全国大范围的干旱对生态领域也造成了较大的影响。“环境质量”“二氧化硫”等热点词的出现,说明网民比较关注冬季供暖燃烧的煤对空气质量造成的影响。

### 7.1.2 生态领域热点专题分析

生态是指生物在一定的自然环境下生存和发展的状态,也指生物的生理特性和生活习性。生态(eco-)一词源于古希腊字,意思是指家或者我们的环境。随着世界工业化的快速发展,如全球变暖、臭氧层破坏、酸雨、资源能源短缺等一系列问题开始曝露出来,这些生态问题已经严重影响人们的正常生活。随着网络普及,网民成为一个庞大的群体,网民可以在虚拟的网络中建立虚拟的关系,可以发表一些现实生活中不敢发表的言论,传统的方法对如此庞大的信息量已不能进行系统分析,对此我们提出了一些有效的方法,既可以及时获取各个地区的生态问题,也可以快速获取网民对于问题的态度,有利于执政者制定措施来快速解决问题。表 7.2 为选取的十大热点专题。

表 7.2 生态领域相关的十大热点专题

热点专题的名称	帖子量/条	总浏览量/条	总回复量/条	参与人数/个	专题热度
计划生育	123	1734185	20885	4863	0.996
转基因育种安全性	98	167908	18364	3246	0.923
雾霾问题	70	146426	10408	2869	0.887
垃圾焚烧	37	477930	7762	2880	0.820
能源枯竭	32	282283	8975	2125	0.784
新能源发展	18	253102	7500	2461	0.741
生物多样性减少	79	374589	6987	1854	0.708
节能减排实施	51	425860	4958	1422	0.685
燃油税上涨	26	260744	3678	1762	0.659
核泄漏污染	44	387615	3984	1201	0.643

从表 7.2 可以看出,生态领域的热点专题中环境问题、资源问题、治理措施问题的热度比较大。同时,根据本书的指标体系定义,对上述热点专题进行热度以及参与度的计算,从计算结果以及对这些问题的分类来看,跟民生相关的生态问题更容易引起大家的讨论。

## 7.2 生态领域意见领袖分析

在生态领域中涉及的用户包括相关帖子的发布者以及参与浏览和回复这些帖子的舆情受众。在信息传播中，一些网民凭借自身表达能力突出、见解深刻等素质，成为政治领域中的“意见领袖”。他们善于从表象信息透析到比较深层的价值内涵和意义，并能对问题进行解析，形成自己的观点并加以表达，他们的观点能够引起众多普通网民的兴趣和参与，促成网络舆论的最终形成。

通过对网络舆情发布者的关注度分析，可以有效地监测出网络舆情在传播与演化过程中具有影响力的意见领袖。本节采用发布者影响力、发布者活跃度 and 发布者价值这三个不同维度的指标对舆情发布者进行分析，采用负面指数、参与度和受众影响力对各个专题的舆情受众进行分析。

### 7.2.1 意见领袖

#### 1) 发布者影响力

计算公式见节 2.3.2。这里， $g_1$  取 0.05， $g_2$  取 0.25， $g_3$  取 0.3， $g_4$  取 0.4，计算出的影响力结果如表 7.3 所示。

表 7.3 生态领域论坛用户影响力排名

用户	浏览数/次	发帖数/条	帖子回复数/条	粉丝数/条	影响力
杨真 M	603285	1	10420	26	0.918
薄荷茶 hoho	2680144	1	7456	25	0.888
自由士 2011	1008064	46	7611	140	0.848
沙滩上的鱼 2014	27950	1	8300	7	0.833
习小兵	134957	1	7431	1	0.796
神完意适	211394	7	6975	76	0.776
说真话好难	2564	2	30	519764	0.719
每天一顿饭	589858	9	5626	409	0.712
野渡自渡人	631463	18	5286	6841	0.698
刘文宪	1834980	1	4000	7704	0.676
只字未提	186560	1	4215	98969	0.665
偶尔来一下	389275	13	4767	608	0.632
宇航摄行者	1003375	1	3973	11	0.604
一文光军	4716874	1	345	6	0.572
龙王视野	305216	25	4025	88	0.560



表 7.3 列出了用户影响力前 15 位的用户。“杨真 M”的影响力最高，同时回复数和浏览次数都较高，可将其视作意见领袖；“自由士 2011”虽然发帖数比较多，但帖子回复数较少，说明他是一个比较活跃而有一定影响力的用户；“说真话好难”这个用户粉丝数很高，说明其影响力确实大，但是他在生态领域的发帖数及帖子回复数显然比较少，说明他只是在其他领域影响力比较大。

## 2) 发布者活跃度

计算公式见节 2.3.2。这里， $g_1$  取 0.6， $g_2$  取 0.4，计算结果如表 7.4 所示。

表 7.4 生态领域论坛用户活跃度排名

用户	发帖数/条	回帖数/条	活跃度
每天一顿饭	9	2742	0.999
黄贤村民	9	2416	0.882
江南 0U	2	2327	0.846
漫步者 chm	1	2047	0.743
无奖瞎猜	0	1837	0.667
反腐倡廉 198	13	1569	0.576
为未来呐喊	3	1562	0.568
实名举报人沈柏松	12	1461	0.537
自由士 2011	46	1385	0.528
徐晖 2014	1	1355	0.492
LOVEGTJA	1	1313	0.477
周丕东	1	1303	0.473
昆仑寓者	4	1202	0.438
琴荷谷的雪	2	1146	0.417
我是张新刚	0	1083	0.393

表 7.4 列出了发布者活跃度前 15 名的用户，表明这些用户在论坛中较为活跃，同时可以作为重点监视的对象。“自由士 2011”和“每天一顿饭”属于活跃性的意见领袖，他们保持着较高的发帖量和回帖量。从该表中还可以看出，用户影响力和用户活跃度并不存在严格的正比关系，影响力高的用户活跃度不一定高。

## 3) 发布者价值

网络中发布者之间会经常互动，发布者的情感价值观也会相互传递、相互

影响。发布者的情感倾向成为了舆情监测的重点。

选择排名前 10 的负向情绪用户（表 7.5），可以看出大多数情感倾向为负值的用户，已经被网站拉黑或者清除，没有被拉黑的用户也可以作为舆情检测的重点对象。

表 7.5 情感倾向负向排名

发布者	情感倾向	发布者	情感倾向
知 8 道 8	-8.39808	柳翊瑜纪	-4.68477
路過再見徒	-8.09808	匿名用户 2	-4.59904
匿名用户 1	-7.89000	匿名用户 3	-4.59904
钱偌媛三	-6.00077	老子就是个黑	-3.99904
xiutu2	-4.68791	牟奎军越	-3.97500

注：爬取数据时不能获取用户名的用户被标注为匿名用户。

7.2.2 舆情受众

舆情受众主要指所有被舆情影响的网络用户。衡量舆情受众主要有三个指标：负面指数、参与度、受众影响力（计算公式同节 2.3.2）。表 7.6 为生态领域的舆情受众指标。

表 7.6 热点专题受众指标

热点专题	负面指数	参与频度	受众影响力
生态环境	0.45874	0.005557	0.042186
生态资源	0.36929	0.007625	0.050528
技术或措施	0.43894	0.008409	0.021936

从表 7.6 可以看出，生态环境专题的负面指数最高，即该专题下的负向情绪占比重较大，可以重点分析；技术或措施专题用户参与度最大，生态资源专题下的中性用户比重比较大，这些用户的观点比较容易受到其他用户的影响。

7.3 专题 1：生态环境

环境问题是由于人类活动作用于周围环境所引起的环境质量变化，以及这种变化对人类的生产和生活造成的影响。人类在改造自然环境和创建

社会环境的过程中，自然环境仍以其固有的自然规律变化着。社会环境一方面受自然环境的制约，也以其固有的规律运动着。人类与环境不断地相互影响和作用，产生了环境问题。

7.3.1 专题发展简介

随着工业化的快速发展，矿物燃料的大量燃烧，温室气体过量排放，全球变暖问题加剧，水资源出现危机，森林锐减，垃圾成灾，物种加速灭绝，生态环境变得越来越恶劣；人们也越来越重视环境问题，认真参与环境问题的讨论；政府也制定了一些有效的方针政策，然而环境问题不是一朝一夕就能改变的，政府不仅要把控全局，而且要认真关注群众生活中的环境问题，积极解决，这样更有利于增强全民改善环境的信心，有利于环境问题的解决。

表 7.7 是生态环境主题的五大大热帖。在帖子的专题传播过程中，大量的网民针对所发表帖子的观点进行了回复，其中也产生了一些新的且具有一定影响力的观点，并引发了网民的关注和讨论。

表 7.7 生态环境专题五大热帖

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
实名举报：广东省吴川市出了一个“损”书记孙茂荣	gody0911	2014/03/23	11485	903	208	1191	85
看铜仁碧江公检法在黄雨岩一案中的表演	我要整死4你	2014/08/18	1023009	1005	210	1256	73
污染厂建在家门口，儿童血铅超标!!	可怜的农民123789	2014/08/02	24324	457	72	532	67
杭州中泰乡垃圾焚烧厂困局(转载)	2014o327	2014/04/28	377665	612	71	738	88
奇闻：深圳为治理垃圾场臭气举行祭天大典，创新环保技术引领全球	兰花萝卜	2014/09/14	66392	504	99	788	82

7.3.2 观点分类与观点列表

通过网络数据信息，针对网民在回复过程中的观点挖掘与倾向进行语义的

处理，形成了本话题内的核心观点。

1) 观点分类

根据观点涉及的特征词汇和领域以及网民发表的各个观点内容本身之间存在的相关性与差异性，可以将所有评论分为以下四个主题。每个主题对应的观点人数比例见图 7.2。

**主题一：有些部门为了一些利益，不惜带头污染环境**

此分类观点包含三个关键特征词：环境污染、部门、利益。对应的观点包括：

- 观点 1. 解决环境问题时，不仅要考虑眼前的利益，还要考虑长远利益。
- 观点 2. 许多人对此感到无奈。
- 观点 3. 持观望态度，对于事情的发展比较感兴趣。

**主题二：垃圾焚烧是垃圾处理好的方式么**

此分类观点包含三个关键特征词：垃圾、处理、焚烧。对应的观点包括：

观点 4. 垃圾焚烧是相对较好的处理方式。

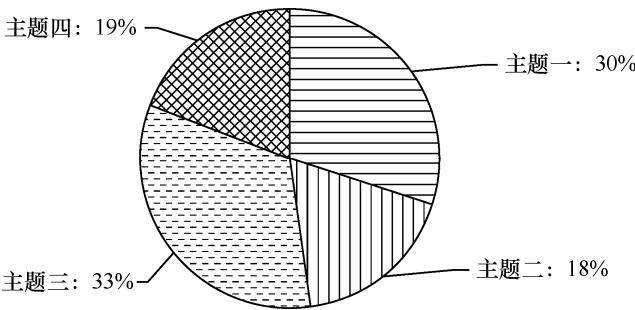


图 7.2 “生态环境”专题观点分类比例

- 观点 5. 学习科学的垃圾处理方式，而不是被动地等垃圾多了再处理。
- 观点 6. 坚决反对在人口密集的地方建造大规模垃圾焚烧厂。

**主题三：合法污染厂建在家门口**

此分类观点包含两个关键特征词：污染、法律。对应的观点包括：

- 观点 7. 和谐社会里我们应该用法律解决问题。
- 观点 8. 这种“合法化”早晚会大白于天下。
- 观点 9. 这种事见多了，希望会有人来处理。

**主题四：祭天治理垃圾场臭气**

此分类观点包含三个关键特征词：祭天、垃圾、治理。对应的观点包括：

观点 10. 环保技术有了划时代的创新。

观点 11. 部分有关部门不作为，百姓民不聊生。

观点 12. 希望能正常呼吸，希望下一代有一个好的环境。

在参与“环境问题”讨论的网民中，网民大都比较关注自身周围的生活环境，积极发言，希望有关部门可以注意到生活环境问题。这说明其实大家都希望有一个好的环境，苦于没有好的解决途径，无奈选择在网络上发声，希望借此引起有关部门的关注，由政府出面治理环境污染从而使生存环境可以得到改善。从一些观点的字里行间中可以看到，网民对于有关部门不作为的一种愤慨。

## 2) 观点分类详细分析

为了更好地从宏观向微观的深入分析，本节对每个分类观点内的网络用户分布以及网络民意进行了进一步的分析，希望通过网络用户真实的反馈信息，来反映他们对这一些问题的思考与建议。

### 主题一：有些部门为了一些利益，不惜带头污染环境

在进一步的数据分析与研究中，对网民在这个主题内的观点进行梳理与分析，发现存在三个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 12%的评论认为解决环境问题时，不仅要考虑眼前的利益，还要考虑长远利益。

例如，“环境的破坏就是永久性的，这样来的利益终究不是长久的利益”。

发展中确实存在许多部门为了暂时利益而破坏生态环境的情况，这类网友评价比较客观。

观点 2. 60%的评论认为这样的事情见多了，已经麻木了。

例如，“也是一种常态了，看上去也还是让人很是无语吧。”

这样的观点具有其存在的理由，在社会快速发展的同时，有许多还不完善的地方，而且一些部门的不良行为很容易传播并造成负面影响，不太乐观的人往往会有这样的想法。

观点 3. 28%的评论持坐等事情的发展态度。

例如，“后续报道快些出来吧，因为我对这个是非常的感兴趣。”

这样的观点具有其存在的理由，因为许多事都是一面之词，许多人会将信将疑，看看有没有其他的真相，然后再去下结论。

综上，在本主题中，具有相应子观点的评论分布比例如图 7.3 所示。

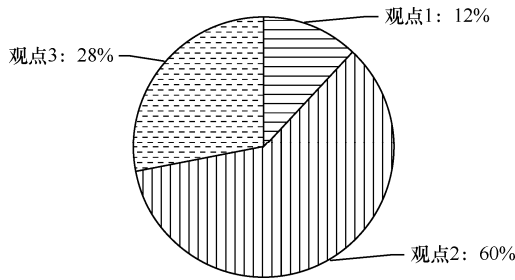


图 7.3 主题一细分子观点比例

### 主题二：垃圾焚烧是垃圾处理好的方式么

在进一步的数据分析与研究中，对网民在这个主题内的观点进行梳理与分析，发现存在三个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 14%的观点认为垃圾焚烧是相对较好的处理方式。

例如，“不建垃圾焚烧厂，垃圾怎么处理？堆你家里去？”

这样的观点具有其存在的理由，如果垃圾不被焚烧，垃圾就会被堆积成山，那样也同样比较恶劣。

观点 2. 31%观点认为学习科学的垃圾处理方式，而不是被动地等垃圾多了再处理。

例如，“应当学习先进的、科学的处理垃圾模式。不可到多得容不下了，才去考虑这些问题，这也算是给有关部门的一个启示。”

这样的观点比较积极，认为做事情要统筹兼顾，不能临时抱佛脚去找应对的办法。

观点 3. 55%观点认为坚决反对在人口密集的地方建造大规模垃圾焚烧厂。

例如，“坚决反对在居住人群密集的城西建如此大规模的垃圾焚烧厂。”

这样的观点比较折中，既希望垃圾可以得到处理，也希望不对大家的生活环境造成较大的影响。

综上，在本主题中，具有相应子观点的评论分布比例如图 7.4 所示。

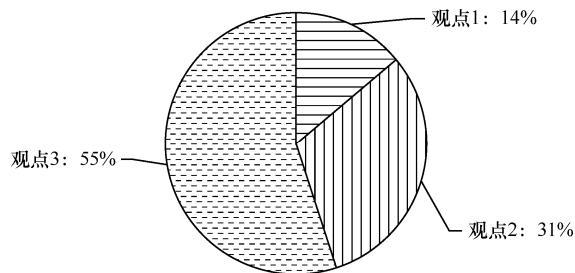


图 7.4 主题二细分子观点比例

### 主题三：合法污染厂建在家门口

观点1. 27%的观点认为可以采用法律解决问题。

例如，“一定要查，这样的事情，不绳之以法，不能大快人心。”“应该用法律的武器维护自己的合法权益。”

这种观点比较积极，身在讲法治的社会主义社会，我们应该运用法律武器，来维护属于我们的合法权益。

观点2. 30%的观点认为这样的事早晚会曝光，坏人一定会得到惩罚。

例如，“相信不久这件事一定引起多方关注的。”“我相信一定是有天理的，中国还是一个法制的国家。”

这种观点比较积极，心中充满正义，认为事情总会往好的方向发展。

观点3. 43%观点认为这种事见多了，希望会有人来处理。

例如，“感觉这个非常不正常啊，到处都是，没有人管的。”“为什么天天都能看到让人上火的事情呢，太多了。”

生活中经常遇到不顺的事，消极情绪很容易积累，这可能是造成这种观点的原因之一。

综上，在本主题中，具有相应于观点的评论分布比例如图7.5所示。

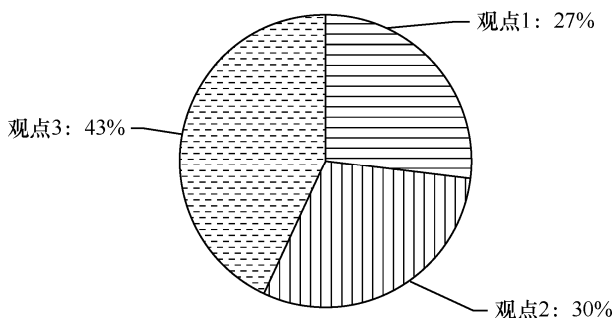


图 7.5 主题三细分子观点比例

### 主题四：祭天治理垃圾场臭气

观点1. 13%的观点认为环保技术有了划时代的创新。

例如，“应该给ZF搬一个最佳创新奖！”“改革开放的排头兵，果然在环保技术上也有了划时代的创新啊！”

这种观点充满嘲讽，在相信科学的时代，竟然想通过祭天的方式来治理环境污染，民众也只能感到惊讶和无奈了。

观点2. 52%的观点认为部分有关部门不作为，百姓民不聊生。

例如，“政府不作为，百姓民不聊生。”“当官不为民做主，不如回家中红薯。”

即使暂时没有好的处理办法，也不能用祭天的方式来糊弄百姓，最后糊弄的只能是自己。

观点 3. 35%的观点希望能正常呼吸，希望下一代有一个好的环境。

例如，“我要正常呼吸。”“可怜的老人和孩子。”“苍天为证，民不聊生，诉求无门，臭不可闻，家家幼子老人长病，天理何在？”“救救孩子，还我健康！关闭下坪垃圾场”。

每个人都希望有一个良好的生活环境，希望给后代营造一个良好的生活环境。

综上，在本主题中，具有相应子观点的评论分布比例如图 7.6 所示。

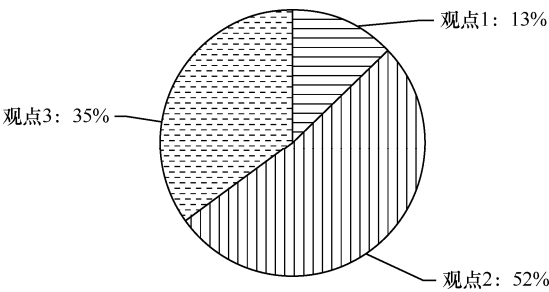


图 7.6 主题四细分子观点比例

### 3) 观点总结

环境问题作为全世界都关注的问题，和我们每个人的生活甚至生存都息息相关。虽然许多网民在 BBS 极力发泄自己的不满，但也说明了他们对于美好生活的向往。大家都生活在同一个环境下，不要被暂时的利益蒙蔽双眼，让未来的我们付出更多的代价。希望有关部门能制定有效措施来保护我们的环境，保卫我们的家园。

### 7.3.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

#### 1) 影响力

专题的传播过程中流量是指相关帖子的回复数量，发展峰值出现在舆论讨论最激烈的时间点。很明显，流量与影响力成正比关系，即流量越大影响力也就越



大。因此,本节分别对新增回复数和总回复数按照时间进行统计,然后对统计结果进行可视化得到波线图(图 7.7,图 7.8),以便发现专题传播的发展峰值。

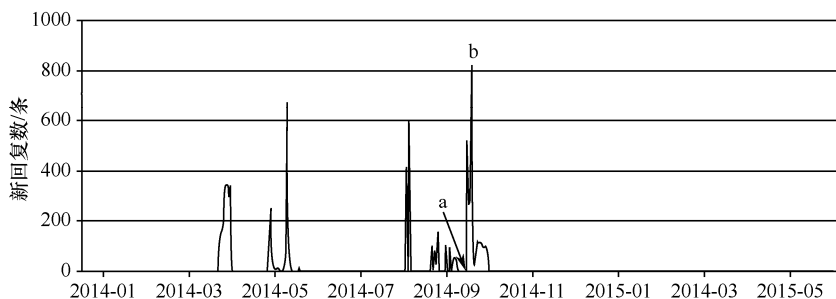


图 7.7 专题日流量时间趋势

点 a: 2014 年 9 月 14 日; 点 b: 2014 年 9 月 18 日

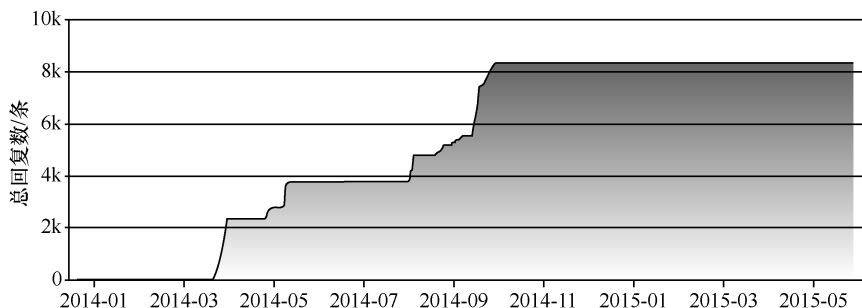


图 7.8 专题总流量时间趋势

从图 7.7 可以看出,在点 a 之前,每天对应的新增回复量有时增加,有时下降,但相对来说,都是比较低,说明这一段时间该事件提出来之后并没有引发很大的影响;而在点 a 之后,每日新增量开始瞬间增高,并呈现持续上涨趋势,在点 b 达到了发展峰值,说明该专题在当点 a 后引发的讨论越来越激烈,在点 b 到达了最激烈的点。在点 b 之后引发的讨论稍微下降,但是偶尔也有上涨趋势,但是总体来说比点 a 之前讨论量增加。在点 b 到达一个发展峰值,可能是在该天发生了一个事件,而引发了人们激烈的讨论。

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库,对专题内的回复内容进行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图(图 7.9)和专题情感分类图(图 7.10)。

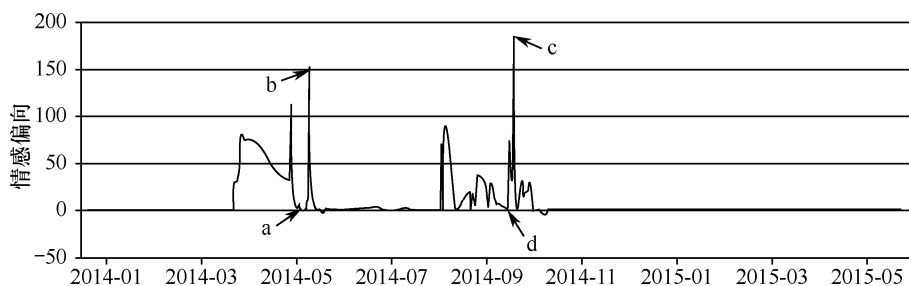


图 7.9 专题情感趋势

点 a：2014 年 5 月 8 日；点 b：2014 年 5 月 10 日；

点 c：2014 年 9 月 18 日；点 d：2014 年 9 月 14 日

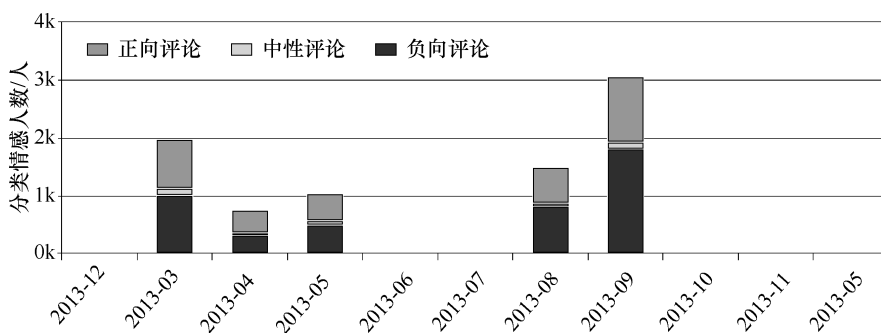


图 7.10 专题情感分类

从图 7.9 可以看出发展峰值。点 b 和点 c 的情感偏向的情感值比较高，说明这两个时间点达到了情感的发展峰值，而在点 a 和点 d 时，情感值较低，随后达到峰值，说明这些点属于情感拐点，必定有一些事情影响了网民的情感。此外，还可以看出在生态领域的网民的情感极易受到影响。结合图 7.9 和图 7.10 可以发现一些规律：影响力发展峰值对应的日期情感值也同时达到峰值，说明流量和情感值也有一定的正相关性。从图 7.10 可以看出，在网民的评论中，负面评论占的比重较大，说明现在网络环境下，网民发言还是比较随意，大部分网民选择在网络上发泄自己的情绪。

### 7.3.4 专题用户情感的分析

本节对各个专题下的人物情感进行了统计分析，由于活跃度对情感的影响较大，所以本节构建了分别以情感值和活跃度为两个坐标轴的专题下任务的散点图（图 7.11）。

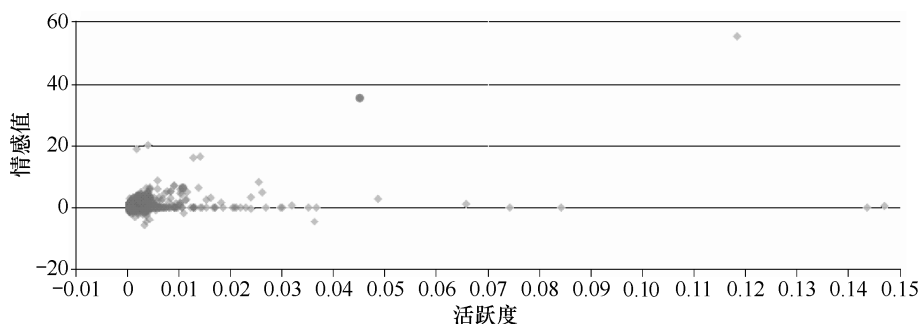


图 7.11 基于活跃度和情感值的用户散点图

圆点代表发帖人，方点代表回帖人

从图 7.11 中可以看出，一般发帖人的情感值都比较高一些。该图中还存在一些活跃度较高但不是发帖者的用户，这些用户有两种可能，一种是该用户作为评论人，评论的内容又引发了一个新的讨论热潮，另一种是属于那些发帖较少，经常参与那些评论或者原帖回复的那一类人，这些人有一定的回复参与量，回复的情感极性通常较低，感情色彩不明显，但是这些人所属领域性比较明显。

为了进一步进行分析，本节将那些情感值负性交大并且活跃度较小的人进行了详细点的描绘，取情感值小于 20，活跃度大于 0.02 的所有用户，作图 7.12。这类用户代表的是那些在论坛活动积极，但是负面情绪较大的人。这些人鉴于他们的活跃度相对来说较大，所以他们影响其他用户的可能性比较大，会给论坛用户灌输一些负面的情绪。因此，很容易从网民中发现这类用户，对其进行权限的限制、屏蔽、删除等操作来避免负面情绪的散播。

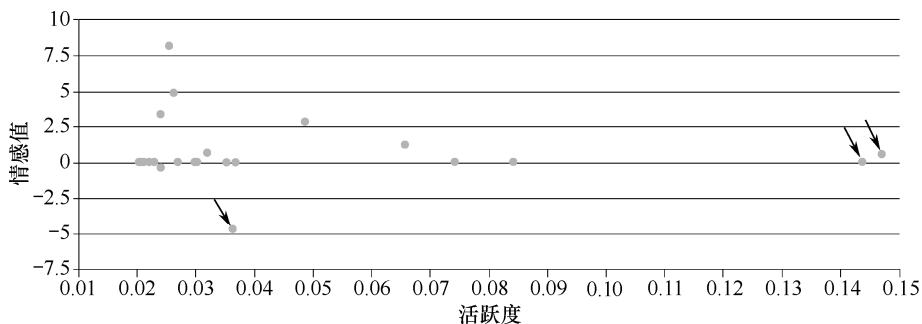


图 7.12 低情感和高活跃度散点图

### 7.3.5 专题网络的构建与分析

构建专题网络的方法见节 4.3.5。构建出生态资源的专题网络,进行可视化,结果见图 7.13。

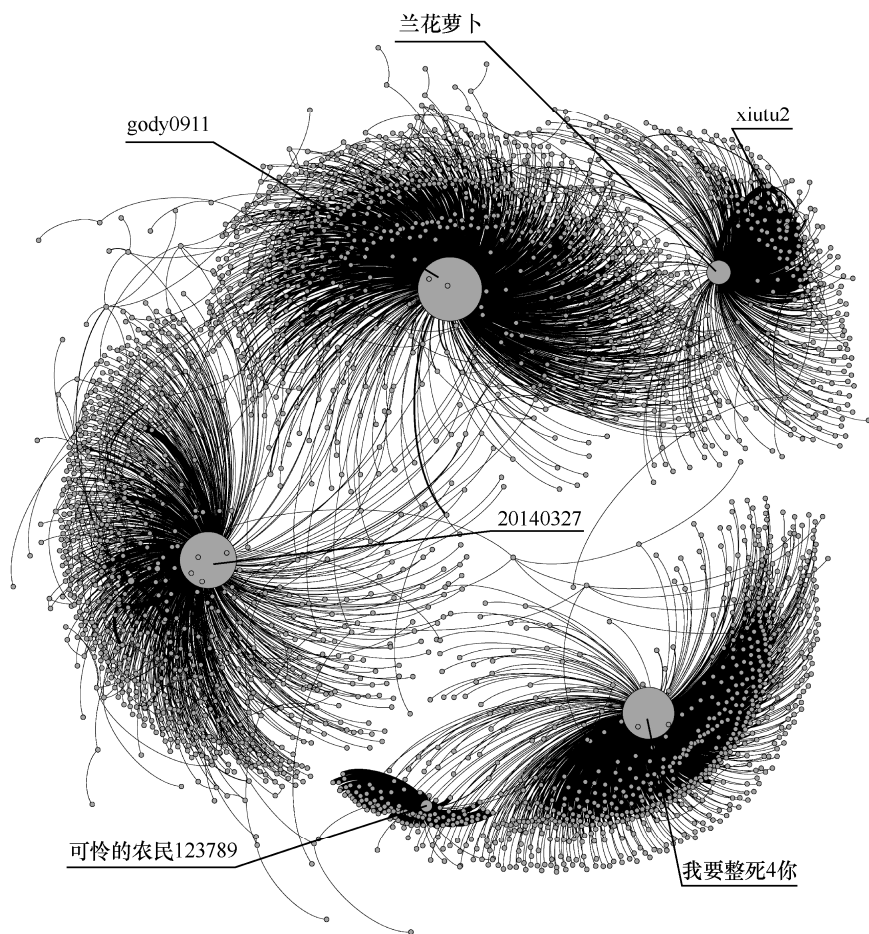


图 7.13 生态环境专题传播网络

由图 7.13 可知,“2014o327”“godly0911”“兰花萝卜”发布的帖子的评论用户之间有交集,其中“2014o327”和“godly0911”发的帖子中有许多小的社区,即用户之间有少量的评论互动。其他两个帖子构成的网络基本上为星形,即其他帖子中用户之间的评论互动较少,主要是以回复帖子作者为主。

1) 网络指标概述

本节分别从社会网络的整体性指标（表 7.8）、中心性与影响力指标、关联性与“小世界”指标进行了分析，以便发现网络中的特殊节点，以便分析个体网络演化与传播。

表 7.8 生态环境专题网络基本指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
2.323	0.001	7	0.780	2.443	0.745

平均度：较小；图密度：该传播网络图比较稀疏。

2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算，这里取点度中心度排名前 6 的节点，列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名，得到对应的数据如表 7.9 所示。

表 7.9 生态环境专题中心性指标

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	排名	绝对值	相对值	排名	绝对值	相对值	排名
我要整死 4 你	2162	0.871	1	0	0	905	5994625	0.047	73
gody0911	2085	0.840	2	0	0	354	5416345	0.052	1
兰花萝卜	1212	0.488	3	12024	0.151	3	7114575	0.040	74
2014o327	1087	0.438	4	56334	0.708	1	5653336	0.050	2
可怜的农民 123789	1012	0.408	5	0	0	1603	7661836	0.037	119
xiutu2	112	0.045	6	1435	0.018	11	7115140	0.040	78

从表 7.9 中可以看出“我要整死 4 你”“gody0911”等 6 个节点的点度中心度排名前 6 位。从中间中心性来看，“2014o327”的中间中心度远远大于其他节点，排在第一位，说明许多节点都需要经过该节点才能建立联系并获取信息。从接近中心性来看，“可怜的农民 123789”和“xiutu2”的接近中心度排名前两位，说明这两个节点与其他节点进行交互比较容易。

图 7.13 的点度中心势指标为 0.87%，该指标值很小，表示该图在度数上向某点集中的趋势很不明显；该图的中间中心势指标为 0.71%，该指标值也比较小，表示该图向某些中间节点聚集的程度比较小，即在整个网络中大部分的节点并不需要经过某些中间节点，就能得到信息。

3) 关联性与“小世界”

图 7.13 的特征途径长度为 2.831，凝聚度为 0.004。该图路径长度的具体分布如表 7.10 所示。

表 7.10 生态环境专题小世界指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	3343	0.045	5	2623	0.035
2	24854	0.335	6	903	0.012
3	31493	0.425	7	27	0.000
4	10907	0.147	8	1	0.000

图 7.13 的聚类系数为 2.233，加权的聚类系数为 0.004。节点聚类系数如表 7.11 所示。

表 7.11 生态环境专题节点聚类系数指标

用户	聚类系数	点对数	用户	聚类系数	点对数
我要整死 4 你	0.000	42556	2014o327	0.000	294528
gody0911	0.000	385881	可怜的农民 23789	0.000	5565
兰花萝卜	0.002	41041	xiutu2	1.250	120

7.3.6 个体网络分析

按照专题领域下的整体网分析之后，选择了一些传播特性相对较强的用户做了个体传播网络图。图 7.14 选取了“2014o327”所发的帖子下对应的所有用户。

1) 结构洞

这里列举节点度数排名前 5 位的用户的结构洞指标，形成相应的数据指标表（表 7.12）。

表 7.12 个体传播网络结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
2014o327	769	675.833	0.879	0.277	0.684
自由灵感 2009	34	27.880	0.820	0.350	0.792
可悲的经租房	20	12.390	0.619	0.588	0.812
谁敢用我马甲	12	10.648	0.887	0.324	0.411
a18621157686	8	3.986	0.498	0.867	0.646

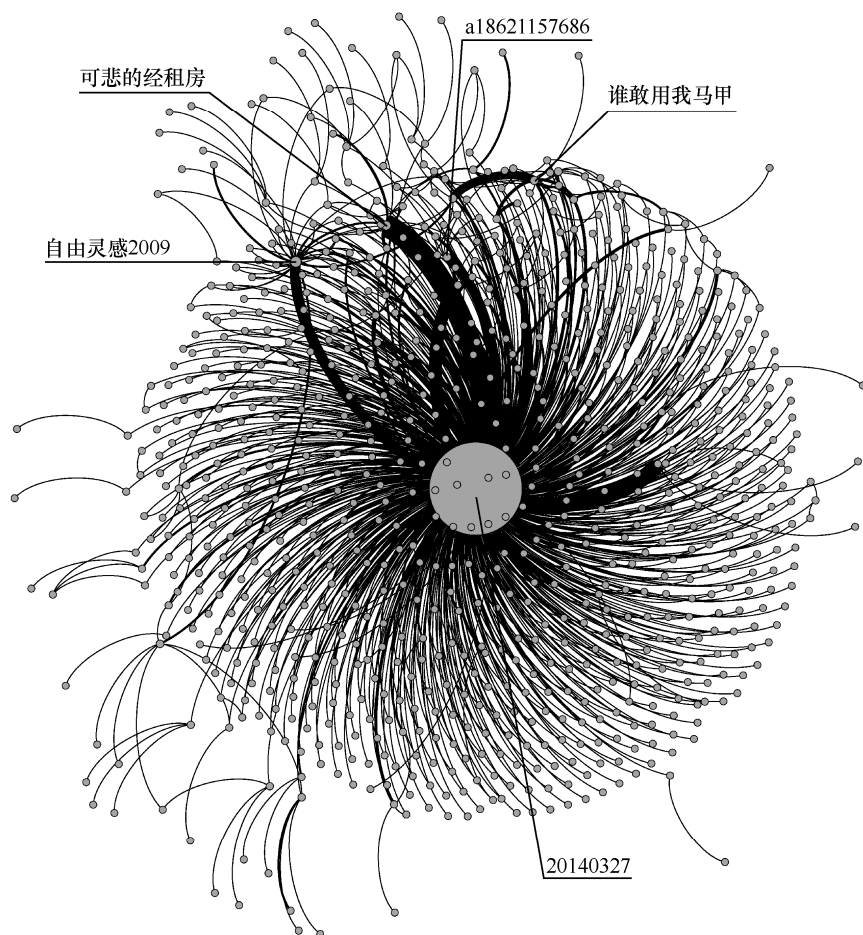


图 7.14 个体网络传播

从表 7.12“有效规模”列看出,“2014o327”所对应的值最大,达到了 675.833,这说明“2014o327”最居于网络的核心,其次是“自由灵感 2009”。“效率”列对应的值最高的为“谁敢用我马甲”,达到了 0.887,说明该用户在社会网络中行动最高效,受到的限制比较小。从“限制度”列看出,“a18621157686”对应的值最大,为 0.867 说明该用户在网络中受到的限制最大。

## 2) 诚实中间人指数

从表 7.13 中可以看出,“2014o327”代表的节点作为中介的次数为 30 次,共联络了 435 对关系,其中 167 对关系属于 HBIO 类,8 对关系属于 HBIO1 类,

4 对关系属于 HBIO2 类, 说明大部分节点属于纯中介, 参与回复主帖的人物之间的关系比较弱; 同时也代表大部分回复是针对主帖回复的, 而只有一小部分回复针对的是评论用户评论的。

表 7.13 个体传播网络诚实中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
2014o327	30	435	167	8	4
自由灵感 2009	5	10	7	1	2
可悲的经租房	4	6	3	0	3
谁敢用我马甲	5	10	7	2	1

HBIO 代表纯中介, HBIO1 代表弱中介, HBIO2 代表非中介。

## 7.4 专题 2: 生态资源

在人类生态系统中, 一切被生物和人类的生存、繁衍和发展所利用的物质、能量、信息、时间和空间, 都可以视为生物和人类的生态资源。地球上有很多丰富的资源, 但可以利用的资源是毕竟是有限的。随着人类的发展、人口激增和生活标准的提高, 增加了对资源的需求, 煤炭、石油、淡水等资源变得越来越少, 人类的可持续发展开始逐渐受到威胁。

### 7.4.1 专题发展简介

随着地球上人口的激增, 生产迅速发展, 水已经变得比以往任何时候都要珍贵。一些河流和湖泊的枯竭, 地下水的耗尽和湿地的消失, 不仅给人类生存带来严重威胁, 而且许多生物也正随着人类生产和生活造成的河流改道、湿地干化和生态环境恶化而灭绝。此外, 世界上资源和能源短缺问题已经在大多数国家甚至全球范围内出现。这种现象的出现, 主要是人类无计划、不合理地大规模开采所至。20 世纪 90 年代初, 全世界消耗能源总数约 100 亿吨标准煤, 预测到 2000 年能源消耗量将翻一番。从石油、煤、水利和核能发展的情况来看, 要满足这种需求量是十分困难的。而且地球上的物种也在加速灭绝, 物种灭绝将给整个地球的食物供给带来威胁, 对人类社会带来的损失和影响是难以预料和挽回的。所以关注地球资源状况, 节约资源等都是需要关注的关键性问题。下面选了五个关于生态资源方面的热帖对其进行分析 (表 7.14)。



表 7.14 生态资源专题热帖

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
藏水北调——人类改造自然、保护环境、促进发展的伟大工程	江南0U	2014/10/23	90764	1439	429	998	89
不支持计划生育的你告诉我中国去哪给你找资源？	swatcaptain	2014/08/16	13176	252	18	279	55
燃油税再次上涨，中国又有希望了！	王梓福	2014/12/13	157176	173	14	222	76
看了后毛骨悚然：中国燃油两次加税的真正原因	bsnhs	2014/12/18	318624	631	47	718	82
说说新疆的石油吧，有些人被骗的很惨！见中线底	德玛西亚希	2014/07/01	193285	289	14	370	40

7.4.2 观点分类与观点列表

通过网络数据信息，针对网民在回复过程中的观点挖掘与倾向进行语义的处理，形成了本专题内的核心观点。

1) 观点分类

根据观点涉及的特征词汇和观点所涉及的领域以及网民发表的各个观点内容本身之间存在的相关性与差异性，可以将所有评论分为以下三个主题。每个主题对应的观点人数比例见图 7.15。

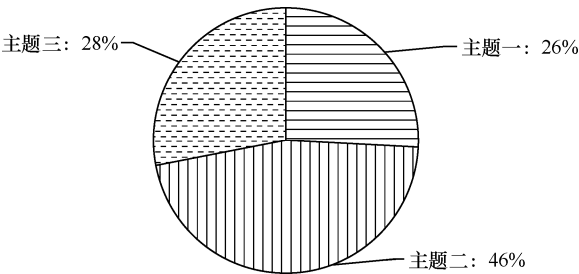


图 7.15 “生态资源”专题观点分类比例

### 主题一：藏水北调工程

观点 1. 藏水北调可行，楼主计划及预算考虑都比较好。

观点 2. 藏水北调可行，楼主计划书有问题。

观点 3. 藏水北调不可行。

### 主题二：计划生育

观点 4. 不反对计划生育，反对民族不平等的计划生育。

观点 5. 反对计划生育，认为生育本身有其规律。

### 主题三：国际燃油降价，中国燃油加税

观点 6. 燃油加税是为了鼓励新能源产业，中国新能源产业已处于世界领先地位。

观点 7. 燃油加税不是为了鼓励新能源产业。

观点 8. 中国新能源不处于世界领先地位。

## 2) 观点分类详细分析

为了更好地对相应的观点分类进行深入的分析，本节对每个分类主题内的网络用户分布以及网络民意进行了分类处理，希望通过对网络用户反馈的真实信息的整理，来反映他们对这一些问题的思考与建议。

### 主题一：藏水北调工程

在进一步的数据分析与研究中，对网民针对“藏水北调工程”主题内的深入观点进行梳理与分析，发现存在三个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 45%的观点认为楼主提出藏水北调工程是个不错的想法，可以实现。例如，“唯楚有才”“比那个渤海地下通道的提议靠谱多了”等。

楼主对自己提出的工程进行了详细的预算及合理性分析，所以占大多非专业的网民并不会发现什么问题，比较赞同。

观点 2. 27%的观点认为楼主不了解工程，预算有很大问题。

例如，“专业的工程量计算不是这样计算的，但工程总造价误差不是很大。”

部分网民了解工程预算，发现楼主的工程预算的问题。

观点 3. 28%的观点认为藏水北调不可行。

例如，“北高南低，调水是个笑话。水质乌黑，即使净化也根本无法饮用。”“痴人说梦”。

一部分网民从其他角度分析，认为这样的工程不合理。

综上，在本主题中，具有相应子观点的评论分布比例如图 7.16 所示。

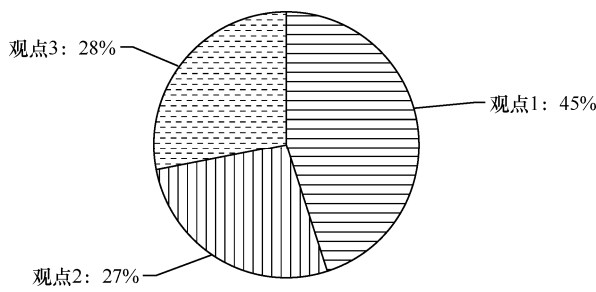


图 7.16 主题一细分子观点比例

主题二：计划生育

观点 1. 72%不反对计划生育，反对民族不平等的计划生育。

例如，“搞计划生育可以，但是能不能对少数民族也实行计划生育，能不能限制非华人向中国移民？广州有三十万黑人，义乌有三十万阿拉伯人，你造吗？”“中国计划生育是没错的，错误的是不平等的计划生育”。

持这样观点的网民也有一定的道理，计划生育比较科学，但是中国许多少数民族并没有实行计划生育，其他民族的人就感觉不公平。

观点 2. 28%反对计划生育，认为生育本身有其规律。

例如，“生育本身有其规律，东亚儒家地区是全世界最最不愿意生孩子的地方。计划的思维本身就是错的，上计划天，下计划地，中间还想计划空气？计划经济也是每年再变，但是计划管制本身就是不对的思维，全世界最最奇葩的思维。”

许多网民认为计划生育本身有规律，或者会受到其他因素的规律影响，人口目前的数量是由于之前不合理计划造成的。

综上，在本主题中，具有相应子观点的评论分布比例如图 7.17 所示。

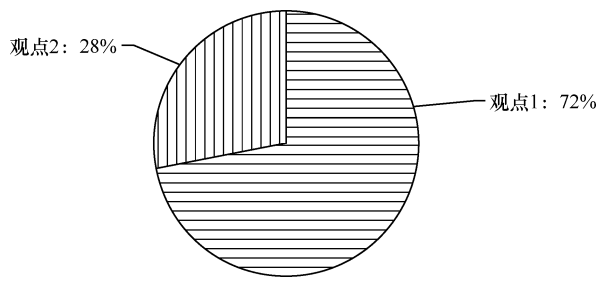


图 7.17 主题二细分子观点比例

**主题三：国际燃油降价，中国燃油加税**

观点 1. 52%的网民支持楼主观点，认为中国燃油加税是为了支持新能源的发展，且中国新能源处于世界领先地位。

例如，“有点道理”“顶下”。

燃油加税在某些方面会促进新能源技术的开发和利用，所以许多网民赞同这样的观点。

观点 2. 33%的网民认为加税不是为了支持新能源发展。

例如，“你的逻辑真是莫名其妙。完全没有条理。”“油价跌了，你加税，没意见；油价涨了，你减税么？不然你这个说法不成立。”

部分网民认为可以从其他方面主动就可以调控来促进新能源的开发和利用，为什么要采用这样被动的措施。

观点 3. 15%的网民认为中国新能源不处于世界领先地位。

例如，“你的逻辑真是莫名其妙。完全没有条理。”“油价跌了，你加税，没意见，油价涨了，你减税么？不然你这个说法不成立”。

许多网民认为楼主并没有拿出来显示中国新能源处于世界领先地位的证据。

综上，在本主题中，具有相应子观点的评论分布比例如图 7.18 所示。

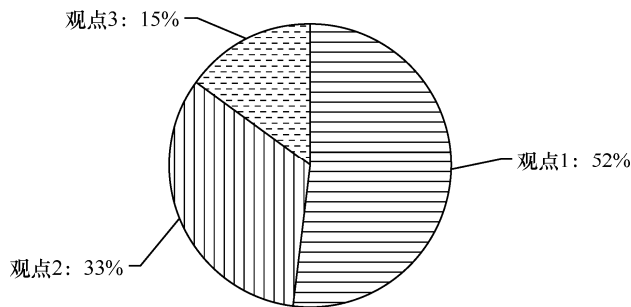


图 7.18 主题一细分子观点比例

**3) 观点总结**

从以上的观点可以看出，淡水、能源、人口等问题都比较贴近生活，网民都认真积极讨论，发表自己的观点，为中国的资源分配出谋划策。有一些好的可行观点可能是解决资源问题的好途径，还有一些网民不求证事实，随意发表负面言论，对于这样的网民应当及时监测出来，对其采取措施，可以为大家营造积极的讨论环境。

### 7.4.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

#### 1) 影响力

专题的传播过程中流量是指相关帖子的回复数量，发展峰值出现在舆论讨论最激烈的时间点。很明显，流量与影响力成正比关系，即流量越大影响力也就越大。因此，本节分别对新增回复数和总回复数按照时间进行统计，然后对统计结果进行可视化得到波线图（图 7.19，图 7.20），以便发现专题传播的发展峰值。

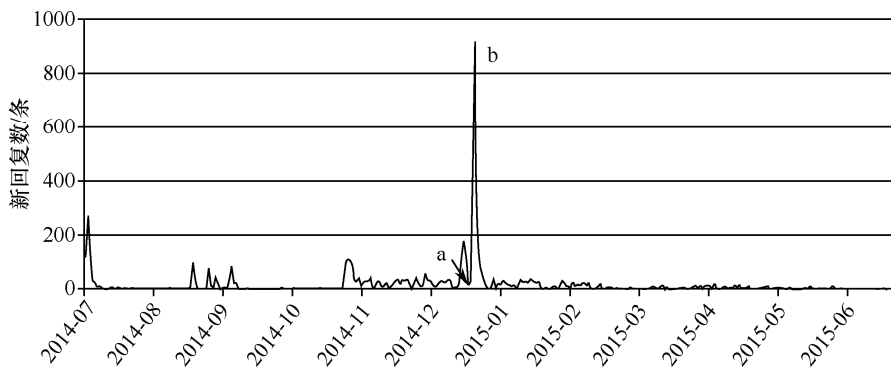


图 7.19 专题日流量时间趋势图

点 a: 2014 年 12 月 17 日; 点 b: 2014 年 12 月 19 日

从图 7.19 可以看出，在点 a 之前，每天对应的新增回复量有时增加，有时下降，但相对来说，都是比较低，说明这一段时间该事件提出来之后并没有引发很大的影响；而在点 a 之后，每日新增量开始瞬间增高，并呈现持续上涨趋势，在点 b 达到了发展峰值，说明该专题在当点 a 后引发的讨论越来越激烈，在点 b 到达了最激烈的点。在点 b 之后引发的讨论稍微下降，但是偶尔也有上涨趋势，但是总体来说比点 a 之前讨论量增加。在 b 到达一个发展峰值，可能是该天发生了一些热点事件，从而引发了人们激烈的讨论。

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库, 对专题内的回复内容进行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图(图 7.21)和专题情感分类图(图 7.22)。

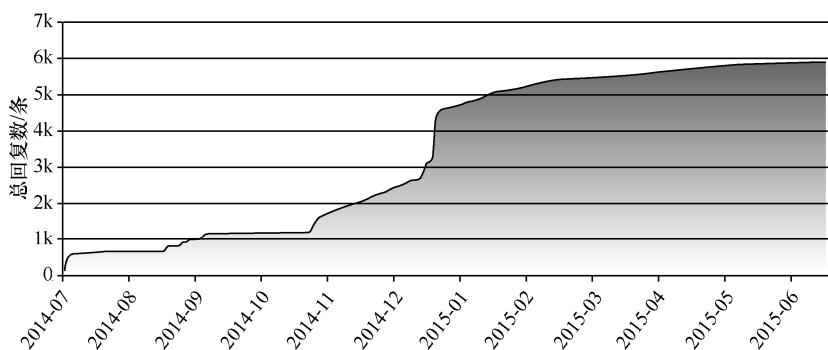


图 7.20 专题总流量时间趋势图

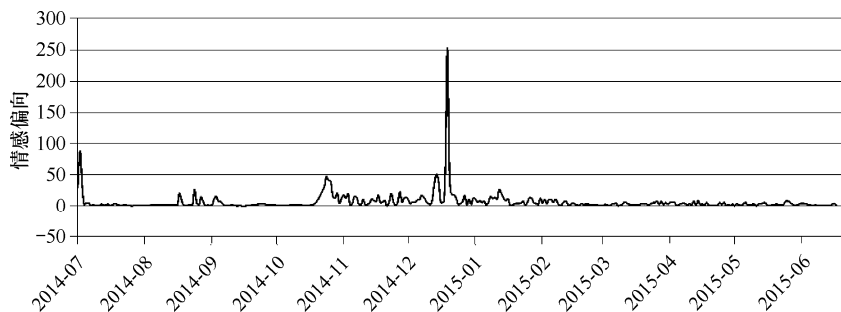


图 7.21 专题情感趋势图

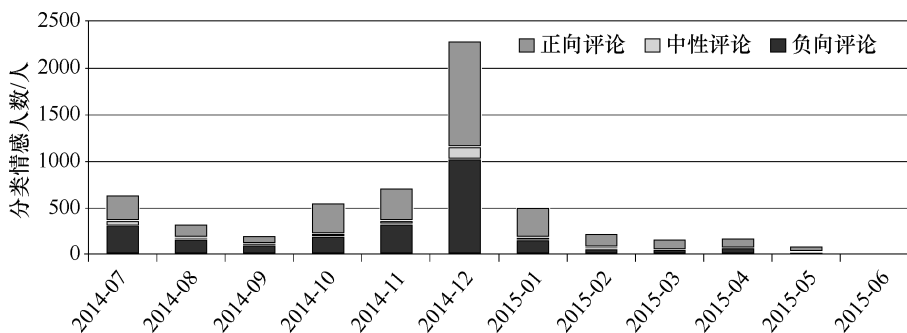


图 7.22 专题情感分类图

从图 7.21 可以看出, 日情感值与日流量曲线呈正相关, 说明对于事件

的评论中正向和负向评论比例比较稳定，即回复量多时情感值就增大，回复量少时情感值就小。此外，从图 7.22 可以看出，在网民的评论中，正向评论普遍比负向评论多，说明在生态资源领域持积极观点的用户存在很大的比例。

#### 7.4.4 专题用户情感的分析

随着近十年来论坛的兴起，网民数量也急剧上升，网民作为网络舆情的一个非常重要因素，行为和言论会引导整个舆论的走向。而在网民中那些发言多，回复多的“意见领袖”或者“领军人物”往往会主导整个舆论的走向。所以说，探索专题下的人物情感可以帮助我们发现该专题下的情感正向用户，同时也可以帮助论坛找出到处散播消极影响的用户，剔除该用户来净化论坛环境。

本节对各个专题下的人物情感进行了统计分析，由于活跃度对情感的影响较大，所以本节构建了分别以情感值和活跃度为两个坐标轴的专题下任务的散点图（图 7.23）。

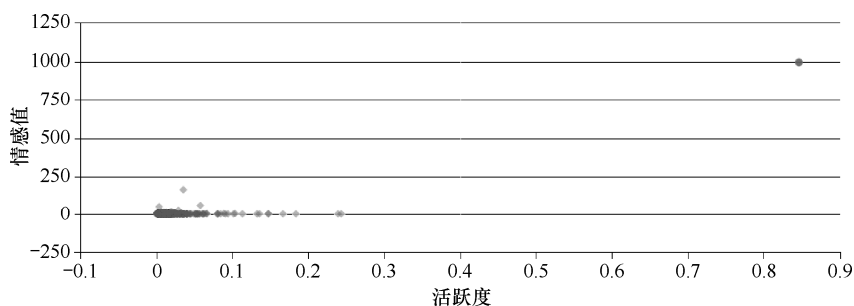


图 7.23 基于活跃度和情感值的用户散点图

圆点代表发帖人，方点代表回帖人

从图 7.23 中可以看出，一般发帖人的情感值都比较高一些，那些圆点的情感值比较高况且活跃度也很大，所以这些用户很可能是意见领袖。但是也存在一些点不是发帖者但是其活跃度较高，说明有两种可能，一种是该用户作为评论人，评论的内容又引发了一个新的讨论热潮，另一种是属于那些发帖较少，经常参与那些评论或者原帖回复的用户，这些人有一定的回复参与量，但是回复的情感极性通常较低，感情色彩不明显，但是这些人所属领域性比较明显。

为了进一步进行分析, 本节将那些情感值负性较大并且活跃度较小的人进行了详细点的描绘, 取情感值小于 20, 活跃度大于 0.02 的所有用户, 作图 7.24。这类用户代表的是那些在论坛活动积极, 但是负面情绪较大的那些人, 这些人的发言或者参与会影响人们对于问题的判断。由于他们的活跃度相对来说较大, 所以他们影响其他用户的可能性比较大, 会给论坛用户灌输一些负面的情绪。因此, 我们很容易从网民中发现这类用户, 并且对其进行权限的限制、屏蔽、删除等操作来避免负面情绪的传播。

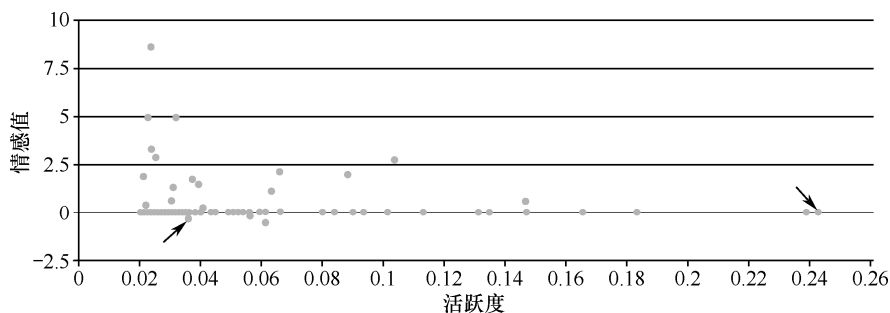


图 7.24 低情感和高活跃度散点图

从图 7.24 箭头所指向的点有两种点比较特殊, 第一种点, 对于本图中的用户来说都是相对活跃度比较大的用户, 因此该用户活跃度比较大, 但是他的情感属于最低点, 这类用户极有可能是负面情绪的发起者, 论坛管理员应该对其采取相应的操作。第二点就是图中标出的另一个点, 该用户活跃度非常高, 比其他用户高出很多, 但是其情感值却紧紧只有 0, 这类用户极易被利用, 发展成为消极情绪的发起者, 应予以关注。

### 7.4.5 专题网络的构建与分析

构建专题网络的方法见节 4.3.5。构建出生态资源的专题网络, 进行可视化, 结果见图 7.25。

由图 7.25 可知, “bsnhs” 和 “swatcantain” 发布的帖子及受众构成了该主题下的两个最大的社区。此外, 还有一些用户发表的帖子形成了一些小的社区。在以 “bsnhs” 为中心点的社区中, 可以在图的右边中看到最典型的一条传播线, “简体中文神” 的用户对原帖进行了评论, 引起了一部分人的关注和讨论。而在以 “swatcantain” 为中心点的社区中, 则较少有度数较大的节点, 这说明没有出现新的话题, 未引起一部分人讨论和关注。



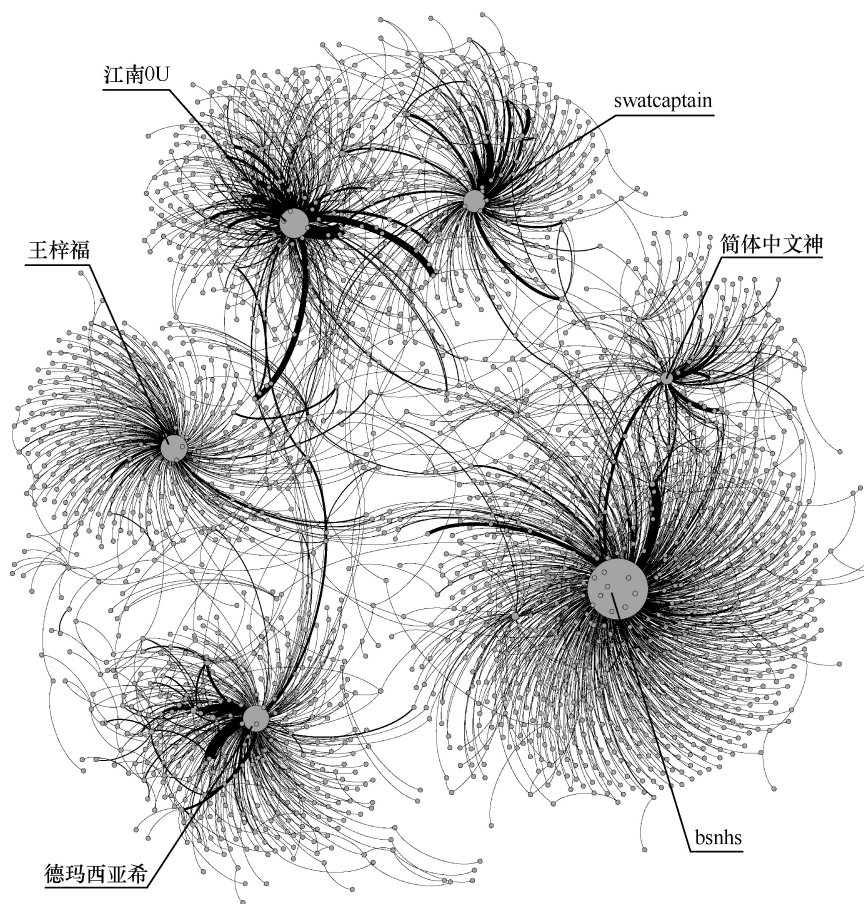


图 7.25 专题传播网络

1) 网络指标概述

本节分别从社会网络的整体性指标（表 7.15）、中心性与影响力指标、关联性与“小世界”指标进行了分析，以便发现网络中的特殊节点，以便分析个体网络演化与传播。

表 7.15 生态资源专题网络基本指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
2.449	0.001	10	0.761	3.850	0.493

平均度：较小；图密度：较小，该传播网络图比较稀疏；网络直径：较长，该网络的传播较远；模块化指标：划分出 7 个社区，且划分的质量较高；平均路径长度：长度较大，网络传播较广；平均聚类系数：较小。

2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算, 这里取点度中心度排名前 5 的节点, 列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名, 得到对应的数据如表 7.16 所示。

表 7.16 生态资源专题中心性指标

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	排名	绝对值	相对值	排名	绝对值	相对值	排名
bsnhs	963	1.5	1	403744	8.240	3	506683	0.437	18
江南 0U	643	1.001	2	615416	12.561	1	1214934	0.182	68
德玛西亚希	422	0.657	3	247639	5.054	6	1216257	0.182	95
swatcaptain	364	0.567	4	0	0.000	349	729717	0.303	2
王梓福	338	0.526	5	0	0.000	616	506683	0.437	1

从表 7.16 中可以看出“bsnhs”“江南 0U”等 5 个用户的点度中心度排名前 5 位。从中间中心性来看, “江南 0U”的中间中心度远远大于其他节点, 排在第一位, 说明许多节点都需要经过该节点才能建立联系并获取信息。从接近中心性来看, “江南 0U”和“德玛西亚希”的接近中心度排名前两位, 说明这两个节点与其他节点进行交互比较容易。

图 7.25 的点度中心势指标为 1.5%, 该指标值很小, 表示该图在度数上向某点集中的趋势很不明显; 该图的中间中心势指标为 12.53%, 该指标值也比较小, 表示该图向某些中间节点聚集的程度比较小, 即在整个网络中大部分的结点并不需要经过某些中间节点, 就能得到信息。

3) 关联性与“小世界”

图 7.25 的特征途径长度为 2.854, 凝聚度为 0.005。该图路径长度的具体分布如表 7.17 所示。

表 7.17 生态资源专题小世界指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	3069	0.004	5	51814	0.068
2	107226	0.140	6	13187	0.172
3	41999	0.055	7	219245	0.286
4	67448	0.088	8	45803	0.060

图 7.25 的聚类系数为 1.100, 节点的聚类系数如表 7.18 所示。

表 7.18 生态资源专题节点聚类系数指标

用户	聚类系数	点对数
bsnhs	0.000	289941
江南 0U	0.000	57630
德玛西亚希	0.002	43660
swatcaptain	0.002	28920
王梓福	0.000	45753

7.4.6 个体网络分析

按照之前整体网的分析，本节从中选取那些节点度数相对较大、传播性比较良好的帖子，做了该帖子下的参与所有用户的个体网络。本节选取的是“bsnhs”发表的一篇帖子，对它进行了指标分析，并且描述其传播路线(图 7.26)。

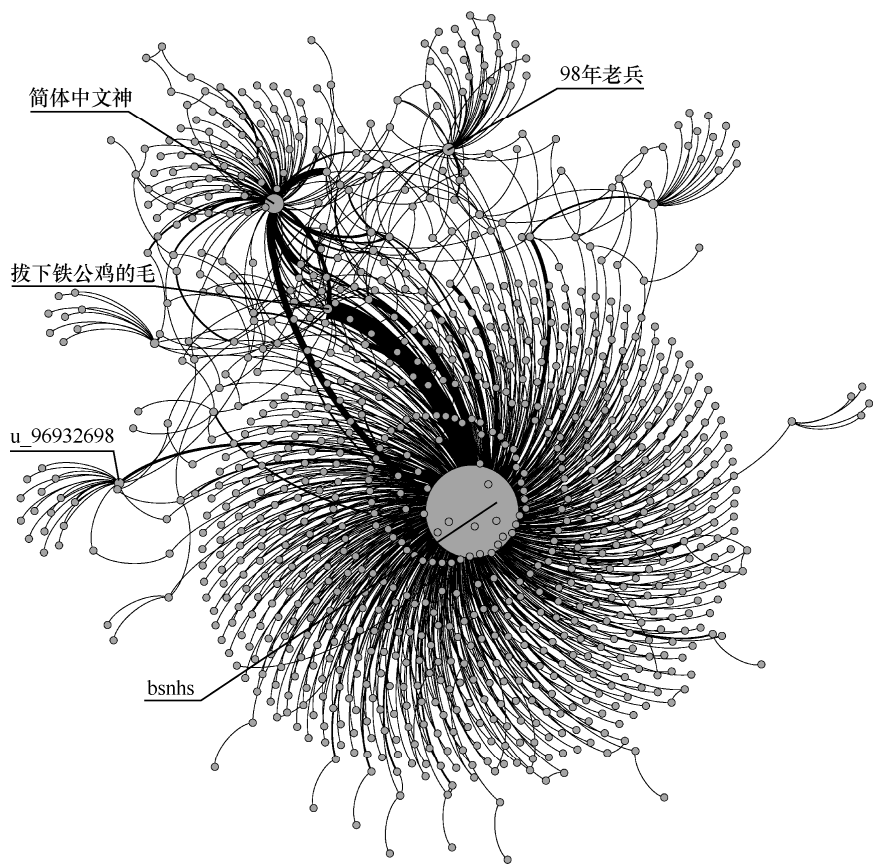


图 7.26 个体传播网络

### 1) 结构洞

这里列举节点度数排名前 5 位的用户的结构洞指标, 形成相应的数据指标表 (表 7.19)。

表 7.19 结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
bsnhs	763	747.263	0.979	0.232	0.977
简体中文神	110	108.454	0.986	0.044	0.503
98 年老兵	43	42.504	0.988	0.055	0.259
u_96932698	21	20.583	0.980	0.082	0.238
拔下铁公鸡的毛	20	13.472	0.674	0.510	0.757

从表 7.19 “有效规模”列看出, “bsnhs”对应的值最大, 达到了 747.263, 这说明“bsnhs”居于网络的核心。“效率”列对应值最高的为“98 年老兵”, 达到了 0.988, 说明该用户在社会网络中和其他节点具有较多的联系, “简体中文神”的限制度值最小, 说明其受到最小的限制。相反, “拔下铁公鸡的毛”的限制度值最大, 说明他们在社会网络中受到较大的限制, 同时他的效率值较小, 说明他可能仅和某一节点联系。

### 2) 诚实中间人

从表 7.20 中可以看出, “bsnhs”代表的节点作为中介的次数为 18 次, 共联络了 153 对关系, 其中 153 对关系属于 HBIO 类。这说明大部分节点属于纯中介, 参与回复主帖的人物之间的关系比较弱。同时也表明大部分回复是针对主帖回复的, 而只有小部分回复针对的是评论用户评论的。

表 7.20 中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
Bsnhs	18	153	153	0	0
简体中文神	16	120	118	2	0
98 年老兵	5	10	10	0	0
拔下铁公鸡的毛	3	3	2	1	0

HBIO 代表纯中介, HBIO1 代表弱中介, HBIO2 代表非中介。

从图 7.25 中可以看出, “bsnhs”发表的帖子后, 大量用户参与的讨论; 在回复的用户中, “简体中文神”发表的评论引起了网民大量回复, 说明该用户在网络传播中可能会是一个意见领袖, 他的意见会影响其他用户的思考和回复。

7.5 专题 3：技术或措施

虽然生态问题显得日益突出，人们的生活环境也受到越来越多的影响，为了改变突出的生态问题，许多研究机构开始研究新的技术，有关部门也开始制定相应的措施，人们开始积极寻求各种策略来应对现有的生态问题。

7.5.1 专题发展简介

生态问题显得日益突出，应对生态问题的技术或措施也纷纷出炉，然而新技术运用或措施实施中可能会出现很多问题，积极找出技术或措施的不足可以对生态问题的解决有很大的提升。为了发掘网民对新技术或措施的看法，选了该专题下的五个热帖（表 7.21），在帖子的专题传播过程中，大量的网民针对帖子所表达的观点进行了回复，其中也产生了一些颇具影响力的新观点，并引发了网民的关注的讨论。

表 7.21 技术或措施专题热帖

帖子标题	发帖人	发帖时间	点击量 /次	正面回 复量/条	中性回 复量/条	负面回 复量/条	作者 情感
中国第一个全开放式有机生态农场即将诞生	琴荷谷的雪	2014/07/08	13816	700	16	429	90
中央文件首提防止“人口过快下降”!	老秦人 1990	2014/01/13	127181	130	18	160	56
人民日报头版——赣南 当振兴成为信念（转载）	shnfeng	2014/08/14	55189	400	30	100	88
无偏见论：转基因是目前最安全的育种方法，转基因食品是未来食品的主流	最后的使徒 000	2014/07/11	4408	182	10	233	74
香港人是高等人，要求中央在珠海助建垃圾岛	一梦轮回	2014/10/26	46160	181	17	211	43

7.5.2 观点分类与观点列表

通过上述的网络数据信息，针对网民在回复过程中的观点挖掘与倾向进行语义的处理，形成了本话题内的核心观点。

### 1) 观点分类

根据观点涉及的特征词汇和领域以及网民发表的观点内容本身之间存在的相关性与差异性，可以将所有评论分为以下三个主题。每个主题对应的观点人数比例见图 7.27。

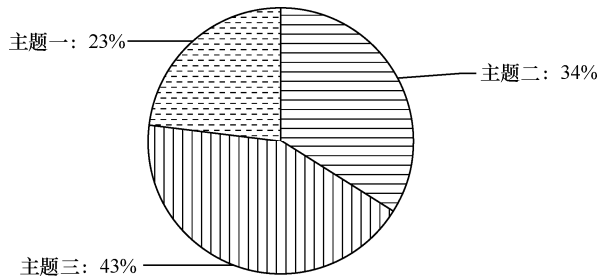


图 7.27 “技术或措施”专题观点分类比例

#### 主题一：防止人口过快下降

观点 1. 支持新政策。

观点 2. 防止过快只是空话，快速下降已是必然。

#### 主题二：转基因育种

观点 3. 转基因不安全。

观点 4. 转基因本身没有问题。

#### 主题三：中国香港要求在珠海建垃圾岛

观点 5. 中国香港弹丸之地，需要多多支持。

观点 6. 中国香港是白眼狼，不用管。

### 2) 观点分类详细分析

为了更好地从宏观向微观的深入分析，本节对每个分类观点内的网络用户分布以及网络民意进行了进一步的分析，希望通过网络用户真实的反馈信息，来反映他们对这一些问题的思考与建议。

#### 主题一：防止人口过快下降

在进一步的数据分析与研究中，对网民在这个主题内的深入观点进行梳理与分析，发现存在两个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 42%的评论认为支持这个提议，认为还可以挽救。

例如，“亡羊补牢，为时不晚。”

部分网民认为适当的鼓励生育政策就可以缓解人口过快下降的趋势，所以

认为还可以挽救。

观点 2. 58%的观点认为为时已晚。

例如,“为时已晚,高物价高房价低福利已经剥削了人们再生的欲望。多生的孩子又将是新时代人们的新三座大山之一。”

随着人们对生活质量要求越来越高,多生孩子会造成经济负担成倍加重,多数家庭不能承受这样的压力,所以多数网民赞同这个观点。

综上,在本主题中,具有相应子观点的评论分布比例如图 7.28 所示。

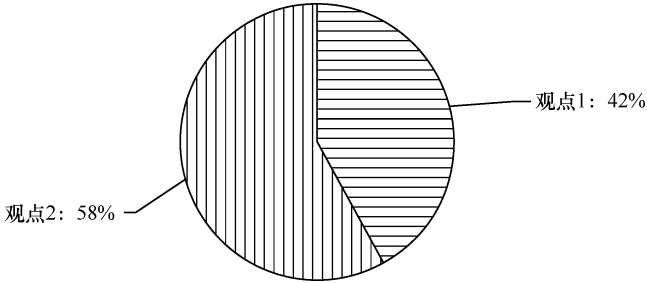


图 7.28 主题一细分子观点比例

主题二：转基因育种

在进一步的数据分析与研究中,对网民在这个主题内的深入观点进行梳理与分析,发现存在两个细分的子观点。这些细分的子观点所反映的内容如下:

观点 1. 77%的评论认为转基因育种不安全。

例如,“又或者你根本就不是国人。作物是在不断进化的,以适应环境,可别忘了人也是慢慢适应的。像你这种人,应该灭九族,凌迟剥皮。”“楼主你是转基因的既得利益者吧?不要在这里洗脑了,要吃你自己吃吧,我们有不是转基因的东西吃就够了,用不拿生命去冒险。”

现有的部分理论认为:导入基因会打乱受体整体基因系统的平衡,在转基因作物体内就会产生新的过敏原毒素及其他致癌物质等。会对食用者造成很大伤害。

观点 2. 23%的观点认为转基因本身没有问题。

例如,“转基因本身没有问题!问题的关键是:控制权在谁手里!白皮猪的转基因食品还真的不能随便吃!谁知道白皮猪们有没有在卖给中国的转基因种子里面植入了特殊的针对中国人有特殊功能的基因片段!楼主是揣着明白装糊涂!要么就是SB,要么就是美狗”。

部分观点认为转基因本身没有问题,只是有些有不法企图的人利用转基因技术来危害目标食用者。

综上,在本主题中,具有相应子观点的评论分布比例如图 7.29 所示。

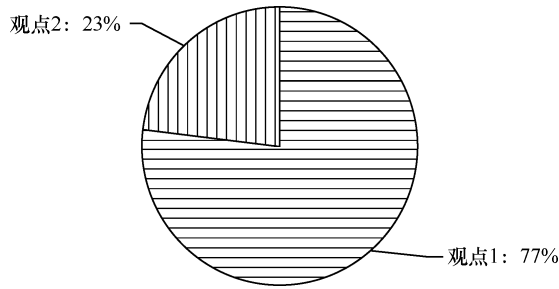


图 7.29 主题二细分子观点比例

### 主题三：中国香港要求在珠海建垃圾岛

在进一步的数据分析与研究中，对网民在这个主题内的深入观点进行梳理与分析，发现存在两个细分的子观点。这些细分的子观点所反映的内容如下：

观点 1. 24%的评论认为中国香港资源比较少，可以支持一下。

例如，“成不成是一回事，可以讨论一下的。中国香港弹丸之地，内地应该多支持。”

香港土地及其他资源都比较少，发展过程中可以对其进行适当的支持。

观点 2. 76%评论认为中国香港比较自私，提无理要求。

例如，“再也不能宠这个没良心的逆子了。”“得寸进尺的无理要求。部分香港人，真的很自私，跟占中一样，总有些混蛋，提出的无理诉求。”

大部分观点认为，香港已经发展很不错了，适当时候应该支持内地发展，而不是向内地索要资源继续自我发展。

综上，在本主题中，具有相应子观点的评论分布比例如图 7.30 所示。

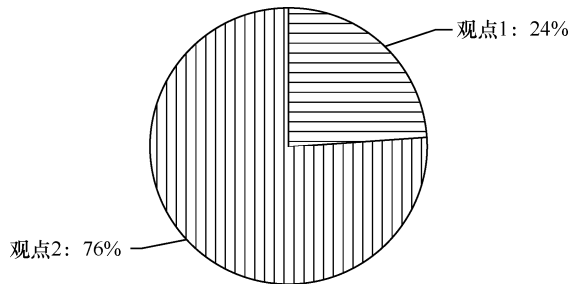


图 7.30 主题三细分子观点比例

### 3) 观点总结

从网友对于技术措施这一领域的评论来看，大部分网友都有自己的看法，也比较关心一些措施及技术对生态造成的影响。但是，大部分网友表达主观意



识较强，话语比较激进，不能客观的而全面的评价一些措施或技术，所以该领域希望专业人士参与来引导网友对于新型未知措施或技术的看法，使大家都可以积极行动起来，加快改变生态现状的进程。

### 7.5.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

#### 1) 影响力

专题的传播过程中流量是指相关帖子的回复数量，发展峰值出现在舆论讨论最激烈的时间点。很明显，流量与影响力成正比关系，即流量越大影响力也就越大。因此，本节分别对新增回复数和总回复数按照时间进行统计，然后对统计结果进行可视化得到波线图（图 7.31，图 7.32），以便发现专题传播的发展峰值。

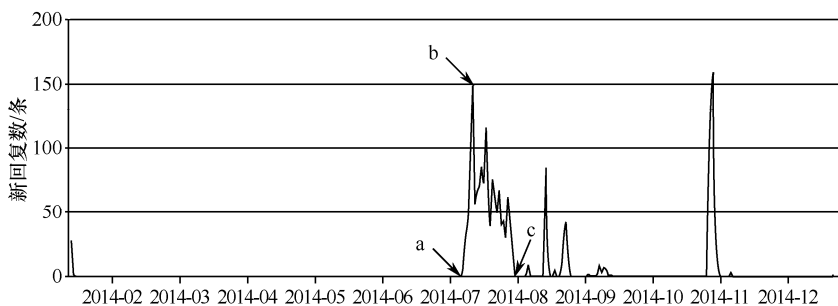


图 7.31 专题日流量时间趋势图

点 a: 2014 年 7 月 7 日; 点 b: 2014 年 7 月 11 日; 点 c: 2014 年 7 月 31 日

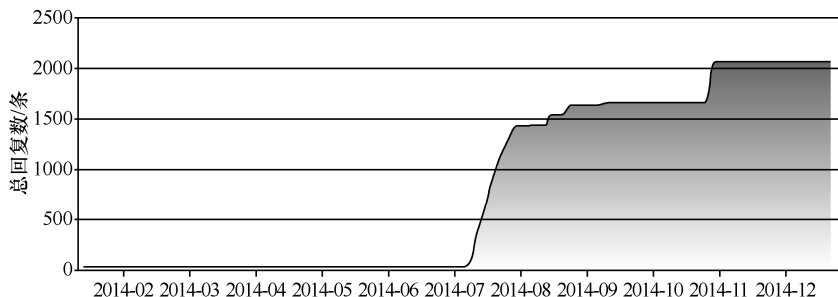


图 7.32 专题总流量时间趋势图

从图 7.31 可以看出，一个事件提出后（在点 a 后）之迅速引起广泛评论，日评论量急剧上升，在点 b 达到了发展峰值；在点 b 之后引发的讨论稍微下降，但是偶尔也有上涨趋势，直到在点 c 回复量达到最小值，说明该事件在 7 月一直是一个热点事件。

2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库，对专题内的回复内容进行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图（图 7.33）和专题情感分类图（图 7.34）。

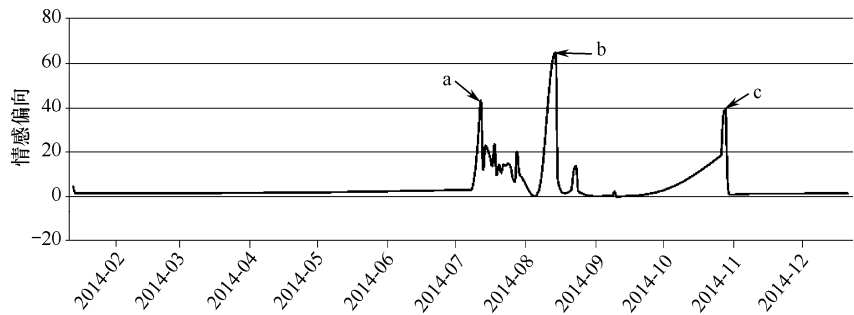


图 7.33 专题情感趋势图

点 a：2014 年 7 月 12 日；点 b：2014 年 8 月 14 日；点 c：2014 年 10 月 28 日

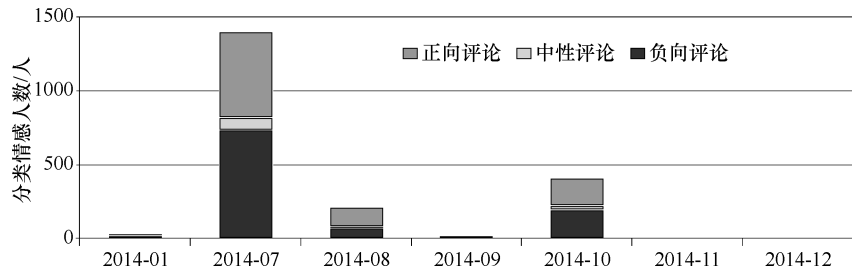


图 7.34 专题情感分类图

从图 7.33 可以明显看出发展峰值和发展拐点。点 a、点 b、点 c 处均为情感峰值，结合日回复量信息发现，日回复量  $c>a>b$ ，而情感值则为  $b>a>c$ ，说明虽然点 a、点 b 回复量相对少一些，但其正向评论值数相对较大，从而情感值比较大。

7.5.4 专题用户情感分析

随着近十年来论坛的兴起，网民数量也急剧上升，网民作为网络舆情的一个非常重要因素，行为和言论会引导整个舆论的走向。而在网民中那些发言多，回

复多的“意见领袖”或者“领军人物”往往会主导整个舆论的走向。所以说，探索专题下的人物情感可以帮助我们发现该专题下的情感正向用户，同时也可以帮助论坛找出到处散播消极影响的网民，剔除该用户来净化论坛环境。

本节对各个专题下的人物情感进行了统计分析，由于活跃度对情感的影响较大，所以本节构建了分别以情感值和活跃度为两个坐标轴的专题下任务的散点图（图 7.35）。

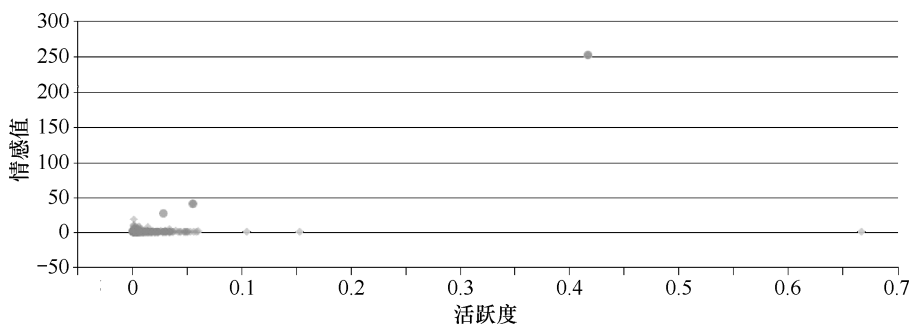


图 7.35 基于活跃度和情感值的用户散点图

圆点代表发帖人，方点代表回帖人

从图 7.35 中可以看出，一般发帖人的情感值都比较高一些，像那些圆点的情感值比较高况且活跃度也很大，所以这些用户很可能是意见领袖。但是也存在一些点不是发帖者但是其活跃度较高，说明有两种可能，一种是该用户作为评论人，评论的内容又引发了一个新的讨论热潮，另一种是属于那些发帖较少，经常参与那些评论或者原帖回复的用户，这些人有一定的回复参与量，但是回复的情感极性通常较低，感情色彩不明显，但是这些人所属领域性比较明显。

为了进一步进行分析，本节将那些情感值负性较大并且活跃度较小的人进行了详细点的描绘，取情感值小于 20，活跃度大于 0.02 的所有用户，作图 7.36。这类用户代表的是那些在论坛活动积极，但是负面情绪较大的那些人，这些人的发言或者参与会影响人们对于生态技术措施的判断。由于他们的活跃度相对来说较大，所以他们影响其他用户的可能性比较大，会给论坛用户灌输一些负面的情绪。因此，很容易从网民发现这类用户，并且对其进行权限的限制、屏蔽、删除等操作来避免负面情绪的散播。

如图 7.36 所示，存在用户活跃度非常高（比其他用户高出很多），但是其情感值却只有 0，这类用户极易被利用，发展成为消极情绪的发起者，应予以关注。

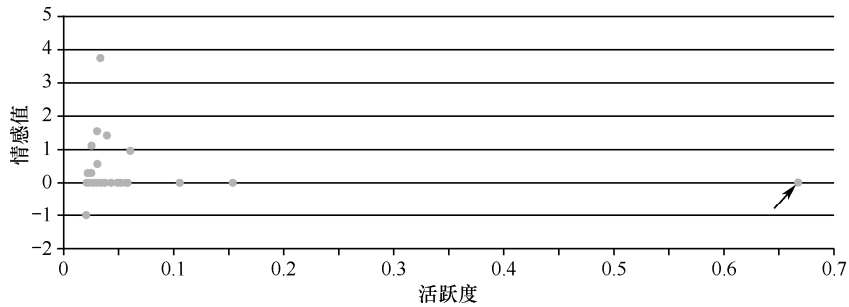


图 7.36 低情感和高活跃度散点图

### 7.5.5 专题网络的构建与分析

构建专题网络的方法见节 4.3.5。构建出生态技术与措施的专题网络，进行可视化，结果见图 7.37。

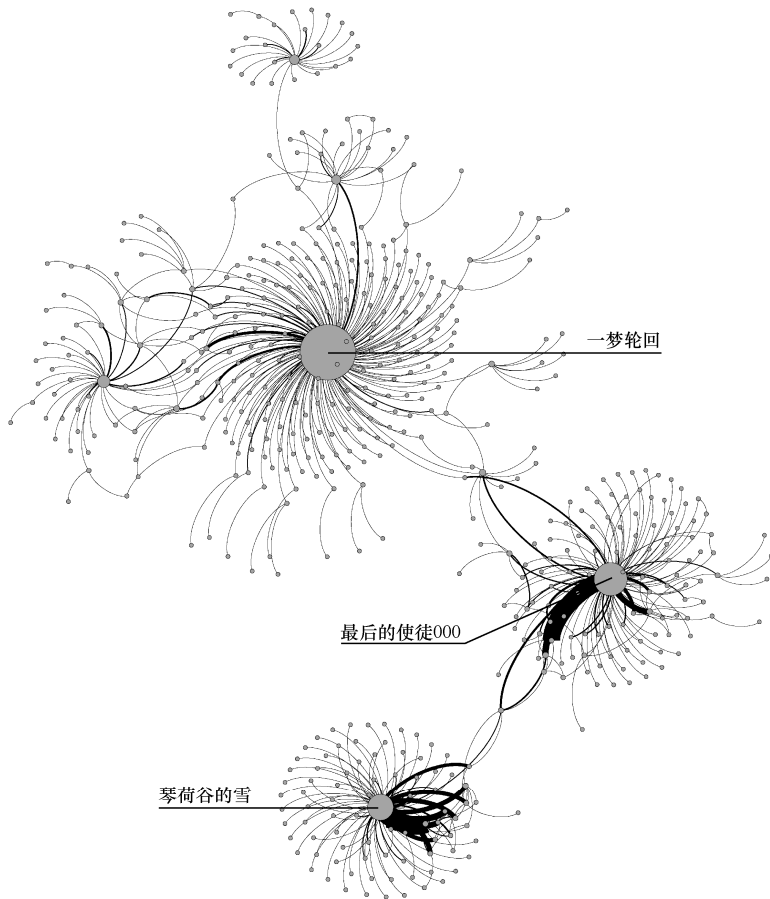


图 7.37 专题传播网络

从图 7.37 中可以明显看到,“一梦轮回”和“最后的信徒 000”的两个用户发表的帖子引起论坛用户广泛的讨论。在整个整体网中,节点之间的联系都比较紧密,可见用户在该主题下讨论的问题都比较集中。除了两个最大的社区之外,“洛开”相对节点度数也较高,他们的观点和意见影响了相当一部分的用户群。

1) 网络指标概述

本节分别从社会网络的整体性指标(表 7.22)、中心性与影响力指标、关联性与“小世界”指标进行了分析,以便发现网络中的特殊节点,以便分析个体网络演化与传播。

表 7.22 网络基本指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
2.365	0.004	10	0.737	4.100	0.475

平均度:较小;图密度:较小,该传播网络图比较稀疏;网络直径:较长,该网络的传播较远;模块化指标:划分出 8 个社区,且划分的质量较高;平均路径长度:长度较大,网络传播较广;平均聚类系数:较小。

2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算,这里取点度中心度排名前 5 的节点,列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名,得到对应的数据如表 7.23 所示。

表 7.23 中心性指标

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	排名	绝对值	相对值	排名	绝对值	相对值	排名
琴荷谷的雪	280	1.140	1	16329	5.254	2	193580	0.288	61
最后的使徒 000	275	1.120	2	18981	6.107	1	193254	0.289	3
一梦轮回	239	0.973	3	0	0.000	276	51432	1.085	1
shnfeng	89	0.362	4	0	0.000	392	291834	0.288	117
王诗瑶	56	0.228	5	8471	2.725	4	193462	0.157	9

从表 7.23 中可以看出“琴荷谷的雪”“最后的使徒 000”等 5 个用户的点度中心度排名前 5 位。从中间中心性来看,“最后的使徒 000”的中间中心度远远大于其他节点,排在第一位,说明许多节点都需要经过该节点才能建立联系并获取信息。从接近中心性来看,“shenfeng”和“琴荷谷的雪”的接近中心度排

名前两位，说明这两个节点与其他节点进行交互比较容易。

图 7.37 的点度中心势指标为 1.13%，该指标值很小，表示该图在度数上向某点集中的趋势很不明显；该图的中间中心势指标为 6.08%，该指标值也比较小，表示该图向某些中间节点聚集的程度比较小，即在整个网络中大部分的结点并不需要经过某些中间节点，就能得到信息。

3) 关联性与“小世界”

图 7.37 的特征途径长度为 3.665，凝聚度为 0.029。该图路径长度的具体分布如表 7.24 所示。

表 7.24 小世界指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	782	0.031	5	1166	0.046
2	9976	0.395	6	5763	0.228
3	2031	0.080	7	1014	0.040
4	4518	0.179	8	29	0.001

图 7.37 的聚类系数为 1.100，节点的聚类系数如表 7.25 所示。

表 7.25 技术或措施专题节点聚类系数指标

用户	聚类系数	点对数
琴荷谷的雪	0.004	3570
最后的使徒 000	0.005	6328
一梦轮回	0.001	19900
shnfeng	0.006	630
王诗瑶	1.100	10

7.5.6 个体网络分析

按照整体网的分析，本节选取那些节点度数相对较大、传播性比较好的帖子，做了该帖子下的所有用户的个体网络。本节选取的是用户“一梦轮回”发表的一篇帖子，对它进行指标的分析，并且描述传播路线。图 7.38 为通过用户发表的帖子的中用户的评论关系构建的传播网络。

1) 结构洞

这里列举节点度数排名前 4 位的用户的结构洞指标，形成相应的数据指标表（表 7.26）。

表 7.26 结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
一梦轮回	200	199.743	0.999	0.009	0.183
洛开	33	32.854	0.996	0.047	0.190
wangyafei55	21	20.630	0.982	0.083	0.214
辣心棒棒糖	8	7.750	0.969	0.164	0.036

从“有效规模”列看出“一梦轮回”对应的值最大，达到了 199.743，这说明用户“一梦轮回”居于网络的核心。“效率”列对应值最高的为“一梦轮回”，达到了 0.999，说明该用户在社会网络中和其他节点具有较多的联系，同时该用户的限制度值最小，说明该用户受到最小的限制。相反，“辣心棒棒糖”和“myjph”的限制度值较大，说明他们在社会网络中受到较大的限制，同时他们的效率值较小，说明他们可能仅和某一节点联系或少量节点有联系（图 7.38）。

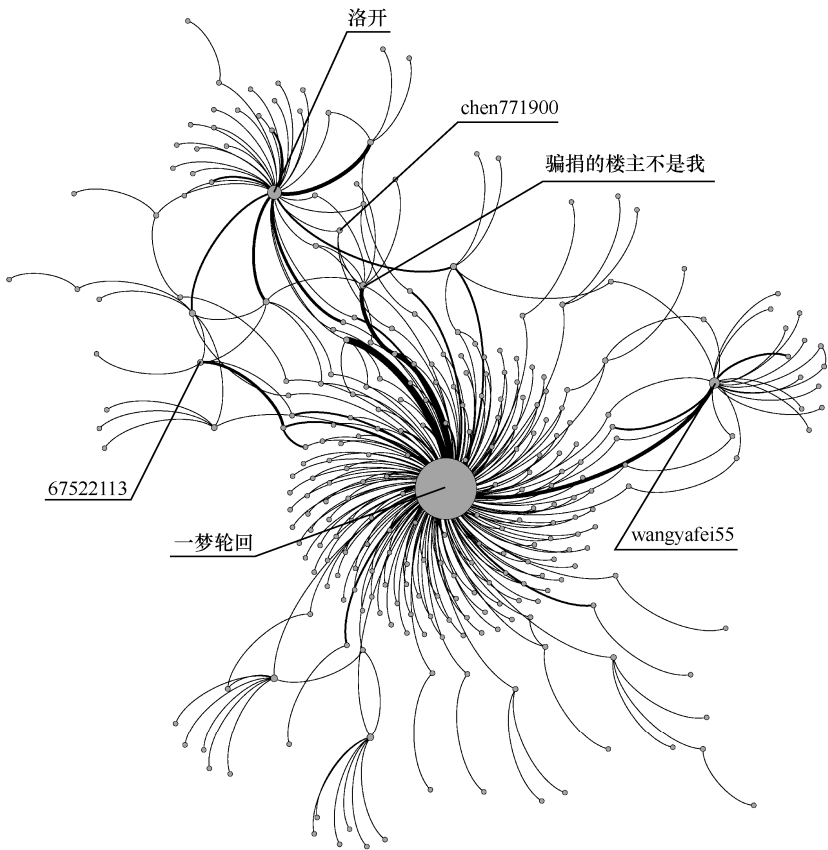


图 7.38 个体传播网络

2) 诚实中间人

从表 7.27 中可以看出,“wangyafei55”代表的节点作为纯中介的次数为 8 次,共联络了 28 对关系,其中 28 对关系属于 HBIO 类。纯中介所占比例最大,同时弱中介和非中介也有一定的比例,表明大部分回复是针对主帖回复的,而其他部分回复针对的是用户的评论。

表 7.27 诚实中间人指标

用户	Size	Pairs	HBIO	HBIO1	HBIO2
wangyafei55	8	28	28	0	0
骗捐的楼主不是我	4	6	6	0	0
67522113	2	1	1	0	0
chen771900	2	1	1	0	0

HBIO 代表纯中介, HBIO1 代表弱中介, HBIO2 代表非中介。

7.6 本章小结

本章主要针对天涯论坛的生态领域的舆情进行了分析,分别从生态环境、生态资源、技术措施三个专题进行了研究和分析。通过该领域整体网的分析,我们发现该领域的帖子数与参与用户数都有了很大提升,说明随着环境问题的日益突出,网民们对改善环境与加强环保的期望与意识越来越高;通过对高频词的时间演化规律的分析,我们发现网民们对于生态问题的关注点跟相应时间点发生的事件有直接的关系;通过对专题传播网络的分析,我们发现网民们对于生态问题的看法与该领域的意见领袖或中间传播者有很大的联系,所以有关部门可以通过对意见领袖或中间传播者进行适当的引导来纠正网民对于生态问题的不合理看法或态度。通过对生态问题分析,我们希望通过有效分析结果可以给有关部门提出一些合理的建议,来加快环境治理的步伐。

文明的生态关系到人民福祉、关乎民族未来的长远大计。因此,面对资源约束趋紧、环境污染严重、生态系统退化的严峻形势,需要把加快生态文明建设作为全面改革的主要任务。为了建立完整的生态文明制度体系,有关部门要对可能形成的环境问题要进行提前预防,同时要加快建立一系列环境保护制度,如源头保护制度、损害赔偿制度、责任追究制度、完善环境治理和生态修复制度等,实现利用制度来对环境污染问题进行管理与控制,这或许是损害群众健康的突出环境问题,并实现生态环境保护目标的有力措施。



## 第8章 国际与军事领域的相关舆情分析

改革开放以来，我国的社会进步与发展步伐加快，国际地位迅速提升，中国开始走向世界，并参与越来越多的国际事务，这无疑涉及两个方面的要求：一是国家利益的拓展，必然要求具有相应的维护利益的军事力量；另一个是作为世界大国，需要承担更多国际义务，其中包括维和、护航、反恐等军事义务。特别是以新技术革命为先导的世界军事变革以及国际情况发生了巨大的变化，国家面临的安全形势也越来越严峻。为了更好地保护改革开放以来我国的建设成果，并保证进一步深化改革过程中维护国内安定团结的环境与氛围，实现保卫国家领土完整并实现国家统一，保障日益扩大的海外利益，建设一支强大的军队对于我国的长治久安具有极其重要的意义。

2014 年国际战略形势的主要特点可以用一个“乱”字概括，包括乌克兰危机继续发酵、俄美欧关系陷入“新冷战”、东海南海局势持续紧张；马航两度发生客机空难、亚航客机失联、“埃博拉病毒”全球肆虐、日本与印尼等多地火山喷发；以及巴基斯坦等国政局持续动荡、奥巴马政府赢了经济却输了中期选举、中国台湾“九合一”选举国民党惨败等，不一而足。国际战略形势之所以出现严重的混乱与失序，主要反映国际格局和世界秩序正处在大变动、大调整过程中，旧秩序尚未退出，新秩序远未建成。预计世界乱局还将持续一个历史时段，并将催生新的国际格局与世界秩序。尽管国际局势如此混乱，但是和平与发展依然是当今世界的主旋律，金砖会议、G7 峰会和 APEC 峰会等国际会议的召开正是对和平发展的践行。

为了客观地分析国际与军事领域的热点舆情，我们从天涯论坛中选取了 2014 年度一些国际与军事领域热点专题的帖子进行深入的总结和分析，本章后续部分将从不同的角度展示该领域舆情分析的结果。

### 8.1 国际与军事领域热点专题和热点词分析

本节选取国际与军事领域的相关帖子进行分析，通过对词频进行统计得到该领域的热点词及演化情况，通过对原帖进行文本聚类得到热点事件，从而确定该领域的热点专题。

### 8.1.1 国际与军事领域热点词分析

热点词出现的频率会随着热点事件的发生而快速增长,从热点词的演化过程可以看出不同时间热点事件的主题变化情况。对2014年全年的数据按季度进行词频统计和排名,结果如表8.1所示。

表 8.1 高频词汇统计

2014.01~2014.03		2014.04~2014.06		2014.07~2014.09		2014.10~2014.12	
词语	词频	词语	词频	词语	词频	词语	词频
事件	36141	事件	27096	事件	20664	事件	12752
马航	15971	马航	11808	马航	9147	亚航	5589
南海	13283	南海	10566	南海	7732	南海	4706
分子	8656	分子	6524	分子	4970	分子	3055
流氓	8577	流氓	6414	流氓	4906	流氓	3037
政府	6506	越南	5120	政府	3795	政府	2269
MH370	5729	政府	4924	美国	3463	美国	2078
黑匣子	5720	美国	4854	黑匣子	3286	大国	2066
越南	4507	黑匣子	4301	昆明	3269	APEC	2021
阴谋	4325	昆明	4280	越南	2779	黑匣子	2021
较量	4318	阴谋	3232	MH17	2498	越南	1621
失事	4305	失事	3211	阴谋	2482	亚航	1620
驾驶员	4291	驾驶员	3209	失事	2470	力量	1608
迷航	4282	迷航	3201	意志	2460	QZ8501	1575
世界	3706	世界	2991	迷航	2452	发生	1564
战略	3638	战略	2848	驾驶员	2449	较量	1561
乌克兰	3199	组织	2316	金砖	2248	阴谋	1543
组织	3006	看到	2294	战略	1971	失事	1522
解体	2987	分析	2267	MH370	1876	驾驶员	1516
民众	2964	峰会	2262	解体	1774	迷航	1515
可能性	2917	解体	2244	系统	1767	战略	1375
结局	2899	MH370	2241	冷战	1742	组织	1192
同盟	2884	冷战	2230	国内	1738	解体	1139
欧美	2875	同盟	2180	民众	1699	西方	1111
赎金	2854	结局	2165	结局	1654	看到	1104

续表

2014.01~2014.03		2014.04~2014.06		2014.07~2014.09		2014.10~2014.12	
词语	词频	词语	词频	词语	词频	词语	词频
俄罗斯	2737	欧美	2140	非盟	1648	冷战	1084
西方	2573	赎金	2136	欧美	1638	国内	1082
政治	1962	强攻	2135	赎金	1632	民众	1079
日本	1852	G7	2135	强攻	1632	系统	1071
社会	1822	西方	1765	乌克兰	1297	俄罗斯	1056
钓鱼岛	1759	日本	1680	西方	1277	金砖	1031
领土	1727	政治	1533	俄罗斯	1257	同盟	1026
武力	1689	钓鱼岛	1471	导弹	1121	欧美	1021
埃博拉	1672	社会	1451	领土	1019	MH17	1021
危机	1656	菲律宾	1433	大国	989	合力	1011
行动	1650	领土	1352	社会	976	赎金	1010
卫星	1617	埃博拉	1348	印度	970	埃博拉	815
雷达	1589	大国	1331	钓鱼岛	969	政治	746
菲律宾	1582	武力	1261	危机	956	日本	725
领导人	1569	舆论	1226	日本	956	社会	719
印度	1552	资源	1219	武力	954	经济	715
美军	1542	美军	1196	埃博拉	953	印度	695
导弹	1541	印度	1196	卫星	950	开放	694
资源	1541	乌克兰	1169	菲律宾	948	领土	691
航空	1532	导弹	1164	媒体	941	钓鱼岛	687
马六甲	1521	领海	1157	资源	890	武力	656
基地	1515	领导人	1154	领导人	890	危机	652

为了更好地观察词频的演化过程,我们用作图工具 NetDraw 对高频词和时间的对应关系进行可视化,得到了高频词演化图(图 8.1)。

从热点词和词频来看,一直在国际与军事领域比较热的词,除了“中国”“美国”“政府”“经济”这些共有的普遍词之外,像“南海”“马航”“事件”“黑匣子”“乌克兰”“钓鱼岛”等这些高频词是 2014 年的热词,一直是热点事件。除此之外,可以从演化图和词频图看出一些热点的演化。例如,在 2014 年 1 月到 3 月“马六甲”“航空”“雷达”等词是热点词,并且只属于这一季度,说明在这段时间内马六甲海峡与航空领域是该季的热点,这对应了 2014 年 3 月 8 号的马航失联事件,3 月引发了激烈的讨论。到了 2014 年 4 月到 6 月“G7”“峰

会”等词成为高频词,说明 G7 峰会引发了激烈的讨论。从 7 月到 9 月,“非盟”这一词出现的频率较大,非盟首脑会议的召开,成为人们谈论的热点。从 10 月到 12 月这三个月,“APEC”“QZ8501”“经济”“开放”等词的出现,说明了在 12 月的亚航失联与 11 月的 APEC 峰会的召开,在网络上引发的大量的讨论。除此之外,像同时归属于四个季度下的词也有很多,说明这些词代表的事件一直是网络中的热点事件,在每个季度下都有相应的帖子或者评论被发布。

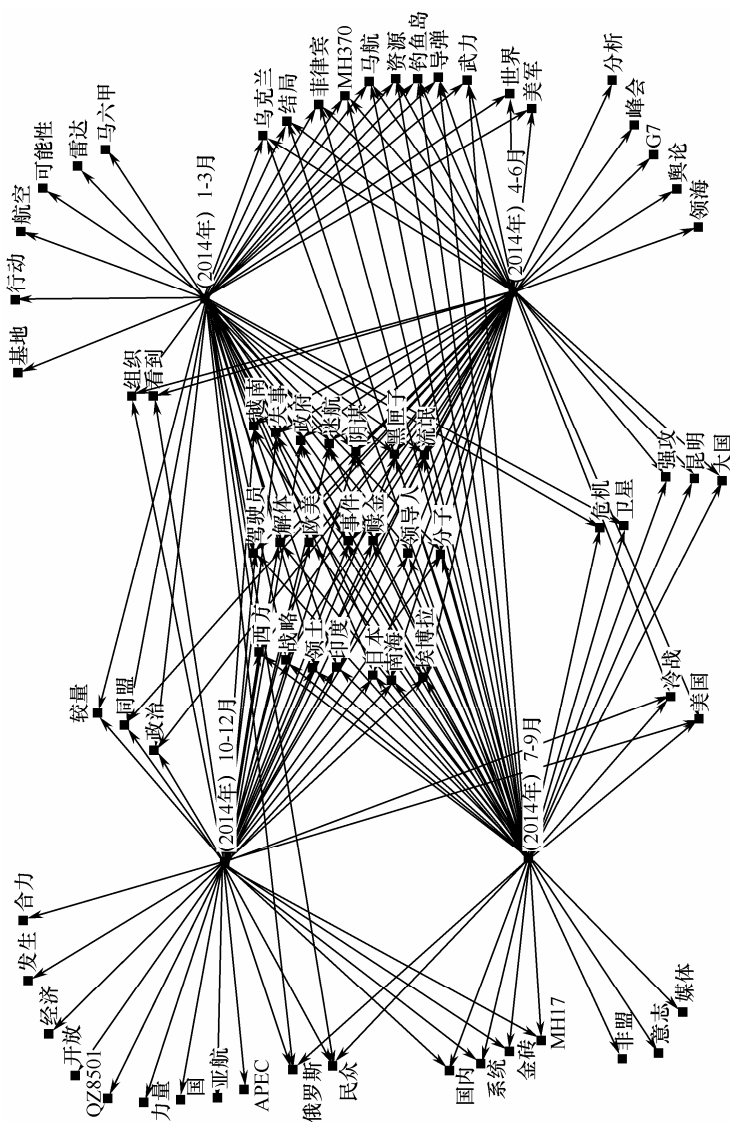


图 8.1 高频词演化图

### 8.1.2 国际与军事领域热点主题分析

第二次世界大战以后，基本上就进入了一个和平的社会，但是这只是表面的，当今社会上仍然动荡与不安。国际上仍然有很多国家和地区遭受着战争带来的痛苦，像“巴以战争”“乌克兰战局动荡”等，所以作为一个倡导和平的中国人，时刻关注国家大事、国际大事是一种民族责任感的表现。

通过对 2014 年国际与军事领域的帖子进行文本聚类得出事件，根据节 2.3.2 的指标进行事件热度的计算，并对这些事件进行摘要的抽取和标题的生成。本节列举热度排名前 10 的热点事件形成表 8.2。

表 8.2 国际与军事领域相关的十大热点专题

热点专题	帖子量/条	总浏览量/条	总回复量/条	参与人数/个	专题热度
马航失联事件	100	73134810	376780	222040	0.425942
钓鱼岛争夺事件	300	40802430	504840	274080	0.251120
巴以和谈再次瓦解	64	544128	8936	5912	0.200090
南海海域争夺	300	31425270	382620	195570	0.191422
G7 峰会的召开	300	2855250	23760	16680	0.188724
非盟首脑会议的召开	300	13150920	117360	73950	0.186178
APEC 峰会在北京召开	300	76947960	217500	132870	0.178453
金砖五国会议首脑会面	300	2829600	34350	17040	0.161998
亚航 QZ8501 的失联	64	1416968	54825	6454	0.063486
博拉病毒肆虐蔓延	49	341964	27956	4945	0.050648

从表 8.2 中可以发现，“马航失联事件”“亚航 QZ8501 的失联”“博拉病毒肆虐蔓延”等事件集中反映了国际事故问题，“钓鱼岛争夺事件”“南海海域争夺”“巴以和谈再次瓦解”等事件集中反映了领土主权问题，“G7 峰会的召开”“金砖五国会议首脑会面”“非盟首脑会议的召开”“APEC 峰会在北京召开”等事件反映了国际上国家之间的合作共荣问题。综上所述，领土与主权、合作共荣和国际事故是国际与军事领域中，网民讨论和关注较多的几个热点专题，本报告以领土与主权、合作共荣和国际事故三个专题为例，对政治领域的网络舆情进行深入的分析。

## 8.2 国际与军事领域意见领袖分析

在国际与军事领域中涉及的用户包括相关帖子的发布者以及参与浏览和回复这些帖子的舆情受众。通过对网络舆情发布者的关注度分析，可以有效地监测出网络舆情在传播与演化过程中具有影响力的意见领袖。本节采用发布者影响力、发布者活跃度和发布者价值这三个不同维度的指标对舆情发布者进行分

析，采用负面指数、参与度和受众影响力对各个专题的舆情受众进行分析。

8.2.1 意见领袖

1) 发布者影响力

计算公式见节 2.3.2。公式中参数， $g_1$  取 0.05、 $g_2$  取 0.25、 $g_3$  取 0.3、 $g_4$  取 0.4，计算出的影响力结果如表 8.3 所示。

表 8.3 国际与军事领域论坛用户影响力排名

用户	粉丝数/个	发帖数/条	帖子回复数/条	回帖数/条	影响力
老百姓有自己的乐	7455	5	24182	89	0.999360300942898
昊天经纬	2018	1	22923	0	0.986306734810743
龙王视野	88	115	16242	913	0.833489924232290
旗扬 2014	401	1	15676	328	0.752276167945017
linnpian	37	1	1020	0	0.697567058947485
frankkee	249	24	4203	128	0.628495296336365
诸葛隆中	269	34	4108	92	0.628248667527961
mrdic	33	1	1789	255	0.601409564191611
寸长 2011	901	45	4468	2	0.587669571353651
二爷的拔火罐 1	762	14	3408	1	0.586773642061063
青狐 123456	32	3	2421	0	0.567896448342214
怀念狼 2013	7	1	967	0	0.555352720741428
二手农民	354	2	4245	1067	0.539327001689286
试试注册一个账号	252	55	6841	4	0.533004577294261

表 8.3 列出了影响力排名前 15 位的用户。“老百姓有自己的乐”与“昊天经纬”的影响力明显最高与其他用户，同时帖子回复数和浏览次数都较高，可将其视作意见领袖。

2) 发布者活跃度

计算公式见节 2.3.2。这里， $g_1$  取 0.6， $g_2$  取 0.4，计算结果如表 8.4 所示。

表 8.4 国际与军事领域论坛用户活跃度排名

用户	发帖数/条	回帖数/条	活跃度
peterdetianya	115	913	1.000000000000000
龙王视野	1	1479	0.733198243836540
二手农民	2	1067	0.722728807835189
道冲不盈	1	1056	0.714285714285712

续表

用户	发帖数/条	回帖数/条	活跃度
13068703619	0	441	0.297872340425531
羽谈飞	5	354	0.244174265450860
刘育新 2013	4	350	0.240459304289091
白庆琳 2012ABC	4	333	0.228976697061803
旗扬 2014	1	328	0.222559945964201
步雨中	0	321	0.216818642350557
tanlienger	0	312	0.210739614994934
1m04349689	1	310	0.210401891252954
林森阳光森林	0	259	0.174940898345153
mrdic	1	255	0.173252279635258
华云之巅	0	238	0.160756501182033

表 8.4 列出了发布者活跃度前 15 名的用户，表明这些用户在论坛中较为活跃，同时可以作为重点监控的对象。“peterdetianya”属于活跃性的意见领袖，保持着较高的发帖量和回帖量。同时，用户影响力和用户活跃度并不存在严格的正比关系，影响力高的用户活跃度不一定高。

3) 发布者价值

网络中发布者之间会经常互动，发布者的情感价值观也会相互传递、相互影响。发布者的情感倾向成为了舆情监测的重点。

表 8.5 列出情感倾向负向程度较大的几个用户。可以看出，该论坛下，有些情感倾向为负值的用户已经被网站拉黑或者清除，没有被拉黑的用户也可以作为舆情监测的重点对象。

表 8.5 国际与军事领域论坛负情感极向用户列表

发布者	情感倾向	发布者	情感倾向
血染的菠菜	-17.1055	我最爱的人是姐姐	-3.711
匿名用户 23	-4.8955	hup2015	-3.6
俯卧撑专业九段 1	-4.56244	大汉冠军侯 806	-3.55851
words2deeds	-3.87392	zjn19691969	-3.4691
匿名用户	-3.82162		

注：爬取数据时不能获取用户名的用户标注为匿名用户。

8.2.2 舆情受众

舆情受众主要指所有被舆情影响到的网络用户。衡量舆情受众主要有三个

指标：负面指数、参与度、受众影响力。（公式具体解释见节 2.3.2）

由表 8.6 可以看出，在国际事故中的负面指数最大，说明在这个专题中，网民的情感倾向普遍偏向负面，体现了网民面对灾难时的同情、惋惜和伤感等负面情绪；而领土与主权专题的参与频度相对较大，说明领土主权问题是网民最关注的热点。

表 8.6 舆情受众指标

热点专题	负面指数	参与频度	受众影响力
领土与主权	0.233920	0.041421	0.026214
合作共赢	0.554520	0.001324	0.049574
国际事故	0.673434	0.012341	0.356234

### 8.3 专题 1：领土与主权

领土与国家存亡有密切关系，因为领土是国家的一部分，形成国家的必要条件，国家行使主权的地域及显示国家独有主权的方式。领土主权是指国家在其领土范围内享有的、最高的、排他的权利，包含所有权和管辖权两方面的内容。领土所有权是指国家对本国领土的占有、使用、收益和处置的权力。基于这种权力，国家可以割让或出租部分领土，并入别国或与别国合并。

领土与主权具有两个含义：第一，国家对于其领土及其中蕴含的自然资源的所有权；第二，国家在其领土范围内享有的领土管辖权。

中国坚定维护国家主权和领土完整，不会受限于“韬光养晦”，不会拿自己的核心利益做交易，不会吞下损害中国主权、安全、发展利益的苦果。维护国家主权与领土完整是每个公民的义务和责任。我们应该维护民族尊严、国家主权统一完整，同破坏祖国统一、民族分裂的行为作斗争，拥护国家的民族政策及外交方针，声讨分裂分子及西方少数国家的暴行。

#### 8.3.1 专题发展简介

领土与主权问题自古至今都是社会的热点话题，与每位公民息息相关。就中国而言，中国与邻国在领土与主权问题仍然存在有关争议。例如，“钓鱼岛问题”“中俄边境问题”“中越海域问题”“中日黄海石油问题”等。国际上也存在很多领土与主权争端引起的问题，如“巴以战争”“伊拉克战争”等。这一系列问题也在论坛上引起了网友广泛且激烈的讨论。

本节将对领土与主权专题相关的舆情进行分析。表 8.7 是该专题下的一些热点帖子以及相关统计指标。



表 8.7 专题热帖

帖子标题	发帖人	发帖时间	点击量 /次	正面回 复量/条	中性回 复量/条	负面回 复量/条	作者 情感
从中国台海到南海： 遏制中国的从不是美国	羽谈飞 白庆琳 2012ABC	2014/07/22	41022	266	556	79	97.475
关于“九段线”两篇 观点截然相反的文章	龙的传人 2013- 2013	2014/06/04	32099	674	201	121	78.257
我爱中国共产党！ (转载)	peterdetianya	2014/10/08	18629	394	373	70	56.889
国际地缘政治分析— 美国俄罗斯中国	罗翊州	2014/02/24	188608	953	651	67	47.065
中国迁都最终方案 大曝光,全国人民一致 拍案叫绝	羽谈飞	2014/11/30	242585	354	923	71	11.720

8.3.2 观点分类与观点列表

以论坛中网民的回复为研究对象，通过观点挖掘和语义分析形成本专题内的核心观点，并对其进行分类和详细描述。

1) 观点分类

根据观点涉及的特征词汇和领域以及网民发表的各个观点内容本身之间存在的相关性与差异性，可以将所有评论分为以下五个主题。

主题一：钓鱼岛的归属问题

此分类观点包含三个关键特征词：钓鱼岛、国土、主权。对应的观点包括：

观点 1. 钓鱼岛自古以来就是中国的领土，归属中国理所应当。

观点 2. 日本之所以敢这样做，因为它的后台是美国。

观点 3. 态度一定要强硬，哪怕是动用武力。

观点 4. 攻打日本，报第二次世界大战的仇。

主题二：巴以冲突之延续不歇的战争

此分类观点包含三个关键特征词：巴以冲突、加沙、袭击。对应的观点包括：

观点 5. 巴以双方相互指责对方破坏停火协议，于是战争继续。

观点 6. 巴以战争的再次发生肯定有美帝国主义的挑拨。

观点 7. 加沙的开火仅仅是个开始，战争会延续到其他地方。

### 主题三：乌克兰局势动荡不安

此分类观点包含五个关键特征词：战局、独立、乌克兰东部地区、克里米亚地区、俄罗斯。对应的观点包括：

观点 8. 乌克兰的战争实际上是美俄大国之间的较量。

观点 9. 克里米亚地区独立是受了俄罗斯的支持。

观点 10. 乌克兰东部地区也会效仿克里米亚地区宣布脱离，乌克兰将会国将不国。

### 主题四：南海问题

此分类观点包含四个关键特征词：石油、海域、选择、问题。对应的观点包括：

观点 11. 南海问题中与中国冲突的国家背后必然有大国的支持。

观点 12. 菲律宾支援朝鲜对抗中国只是虚张声势。

观点 13. 中国现在强大了，不是之前的中国，不要对这些小势力屈服，不惜动用武力。

### 主题五：伊拉克遗留的战争问题

此分类观点包含四个关键特征词：侵略、内战、ISIS、反叛。对应的观点包括：

观点 14. 伊拉克自从 2003 年被美国攻占以来，一直战乱不断，说明主权的重要性。

观点 15. ISIS 就是美国养的一个“儿子”，美国不自己出面，让 ISIS 来搅乱国际治安。

观点 16. 这是美国一种变相的侵略他国、霸占资源的手段，将来还会有其他的国家像伊拉克这样被以“正义”之名侵占。

每个主题对应的观点人数比例见图 8.2。

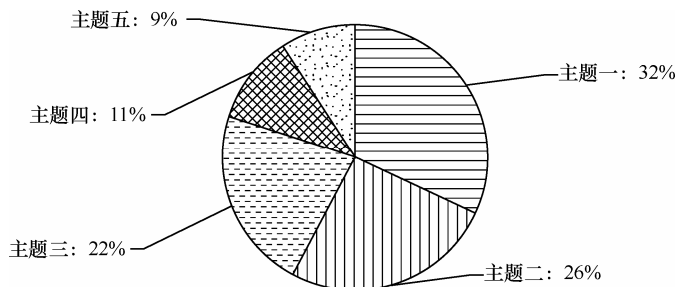


图 8.2 “领土主权”专题观点分类比例

从图 8.2 可以看出,在参与“领土与主权专题”讨论的网民中,大部分网民关心的是与中国有关的主权问题,如“钓鱼岛问题”“南海问题”等。这说明无论在现实或在虚拟网络中,都能看出作为一个中国人应有的爱国热情,和对美帝国主义劣迹行径的一种唾弃和愤怒。同时,有些言论也不是很恰当,类似中国已经强大起来了,动用武力解决问题等。这些言论不符合中国以“和平”为主的外交政策,中国在国际中上扮演了一个和平者的形象,能用和谈的方式解决的问题,绝不使用武力。但是,也请网民认识到这不是懦弱,对于特定问题,如果和谈解决不了,我国也不排除使用武力。

## 2) 观点分类详细分析

为了更好地从宏观向微观的深入分析,本节对每个分类观点内的网络用户分布以及网络民意进行了进一步的分析,希望通过网络用户真实的反馈信息来反映他们对这些问题的思考与建议。

### 主题一: 钓鱼岛的归属问题

在进一步的数据分析与研究中,对网民在这个主题内的深入观点进行梳理与分析,发现存在 4 个细分的子观点。这些细分的子观点所反映的内容如下:

观点 1. 58% 的评论认为钓鱼岛自古以来就是中国的领土,归属中国理所当然,那些妄想和中国争夺钓鱼岛的国家,不论是谁都不会得逞的,我国在钓鱼岛问题上的立场很明确。

例如,用户“lianglong1988”评论中说过:“钓鱼岛及其附属岛屿是中国领土不可分割的一部分。无论从历史、地理还是从法理的角度来看,钓鱼岛都是中国的固有领土,中国对其拥有无可争辩的主权。中国坚决反对和遏制日本采取任何方式侵犯中国对钓鱼岛的主权。中国在钓鱼岛问题上的立场是明确的、一贯的,维护国家主权和领土完整的意志坚定不移……”

观点 2. 24% 的评论认为日本之所以敢这样做,因为它的后台是美国,美国为了遏制中国的发展,怕中国的快速发展会威胁美国霸主的地位。

例如,用户“破坏神 2013”说过:“美国处处煽风点火,拿钓鱼岛来说,将其纳入美日安保条约,完全忽略中国领导人邓小平与日本当年达成的停火协议,这不是点火是干么? 中国和日本怎能平静,威胁这个区域安全的,相反是美国。如果中国试图限制本区域的航行自由或忽视国际规则,美国将会寻求其他方法来解决……”

观点 3. 13% 的评论认为我国态度一定要强硬,哪怕是动用武力,中国虽然是坚持和平的外交政策,但是在国土完整上确实立场坚定。

例如, 用户“狂妄的毛驴”评论说: “对于日本解决钓鱼岛问题所发表的、动用武力解决钓鱼岛问题的说法, 我们也不能一忍再忍, 在和平解决不能解决问题的情况下, 我们一定要态度强硬, 不排除使用武力……”

观点 4. 5% 的评论认为攻打日本, 报第二次世界大战的仇。显然这个观点不是太合适, 虽然中国现在已经强大了, 但是这不符合我国的和平外交政策。

例如, 用户“江城五月数落梅”评论说: “在 21 世纪, 中国已经变得十分强大了, 去年超过了美国, 成为第二大国, 我们绝对有能力和信心来保卫我们国家的领土, 最好的办法, 就是攻打日本, 使小日本成为中国殖民地, 钓鱼岛自然就是我们的……”

综上, 在本主题中, 具有相应子观点的评论分布比例如图 8.3 所示。

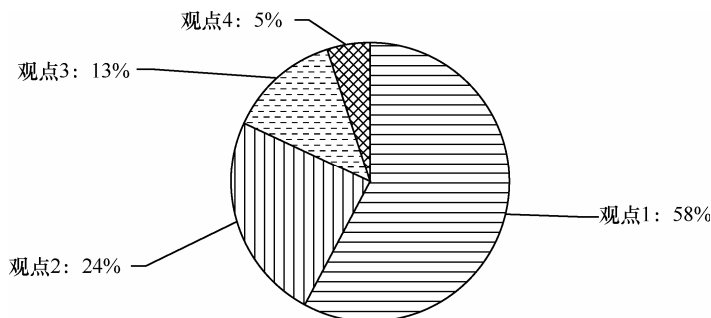


图 8.3 主题一细分子观点比例

## 主题二：巴以冲突之延续不歇的战争

在进一步的数据分析与研究中, 对网民在这个主题内的深入观点进行梳理与分析, 发现存在 3 个细分的子观点。这些细分的子观点所反映的内容如下:

观点 1. 45% 的评论认为巴以双方相互指责对方破坏停火协议, 导致了战争的持续。

例如, 用户“yxtgh”评论说: “双方发生战争的基本原因就是因为他们一个地方的开火, 相互推卸责任, 指责对方没有遵守停火协议, 故意挑起战争的事端……”

观点 2. 34% 的评论表示巴以战争的再次发生肯定有帝国主义的挑拨。常年的战争耗费大量的军费, 没有帝国主义的支持, 也不可能持续那么久, 所以说战争受益的是那些售卖军火的强国大国。

例如, 用户“江湖之远 CQ2012”评论说: “在叙利亚问题是‘人权’大于‘主权’, 在乌克兰问题上‘主权’大于‘人权’。但这一次, 无论巴勒斯坦人的‘主

权’还是‘人权’，都大不过美国与以色列的‘友好’关系。就在7月25日，联合国提起了谴责以色列屠杀巴勒斯坦平民的动议，只有美国、这个以捍卫世界和平与人类正义自我标榜的国家，投了唯一的一张反对票。这就是美国捍卫的‘正义’。在巴以问题上，美国自私邪恶、践踏人类正义、反人类的本质暴露无遗”。

观点3. 21%的评论认为加沙的开火仅仅是个开始，战争会延续到其他地方。

例如，用户“巴国白鹤”评论说：“现在的问题是以色列不接受埃及提出的协议，巴勒斯坦方面接受。我们没有任何条件，我们很想讨论如何稳定局势，建立一个长期的和平环境。而以色列年复一年地重复对加沙地带发动袭击。加沙已经被封锁了7年，我们想解除封锁，让人员和物资在这里自由流通，重建被以色列摧毁的基础设施，重建机场，修建海港，这些都是在和以色列谈判时非常重要的因素。这些算是条件吗？不是！这些都是在联合国、欧盟、美国、俄罗斯见证下，我们和以色列政府已经签订的协议中的内容。”

综上，在本主题中，具有相应子观点的评论分布比例如图8.4所示。

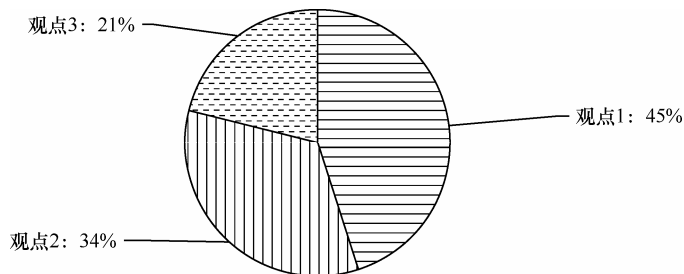


图 8.4 主题二细分子观点比例

### 主题三：乌克兰局势动荡不安

在进一步的数据分析与研究中，对网民在这个主题内的深入观点进行梳理与分析，发现存在3个细分的子观点。这些细分的子观点所反映的内容如下：

观点1. 32%的评论认为乌克兰的战争实际上是美俄大国之间的较量，乌克兰东部地区靠近俄罗斯，得到俄罗斯支持，其他地方得到美国支持。

例如，用户“iluvchina”评论说：“欧盟每打一张牌都以保护自己的利益为底线。所以，乌克兰闹到今天这样，乱得不可开交，最好分裂，他们的目的就达到了。前有叶利钦，后有萨卡什维利，现在季莫申科也逃不出这个老套路，

亲西方对政客有益，对国家没有好结果，都是欧美的傀儡，平时被用来对付俄罗斯，关键时刻从来都是被抛弃，很正常……”

观点 2. 51% 的评论认为克里米亚地区独立是受了俄罗斯的支持，得到了俄罗斯的出兵相救。

例如，用户“战神之武士”评论道：“俄罗斯议会同意克里米亚地区加入俄罗斯，俄罗斯总统普京在协议上签字。普京在随后的讲话中，感谢了几个国家，在俄罗斯与西方的争斗中，理解俄罗斯的行为，其中就点名感谢了中国。感到孤独的俄罗斯和普京，这是第二次拉中国垫背了。第一次是在 2014 年 3 月 3 日，俄罗斯外交部发表声明说，俄罗斯和中国对乌克兰局势的看法‘广泛一致’。因为 3 月 3 日，俄罗斯外长拉夫罗夫和中国外长王毅，进行了电话交谈，两国外长就乌克兰局势交换了看法……”

观点 3. 17% 的评论认为乌克兰东部地区也会效仿克里米亚地区宣布脱离，乌克兰将会国将不国。一次美俄之间的掰手腕，受害的还是乌克兰人民。

例如，用户“alanchanchn”评论道：“克里米亚地区乃至整个乌克兰东部地区会不会并入俄罗斯，在当前的国际大棋局中，并不具有决定性意义。战略对决的地点已经很明白，是中国的台湾、南海、东海。在远离主战场的地方取得一点胜利还是失败，其实无关紧要。美国仍然可以选择尽快取消对俄罗斯的制裁，承认克里米亚地区的现实，取消对乌克兰反对派的支持，允许普京建立一个亲俄的乌克兰政府……”

综上，在本主题中，具有相应子观点的评论分布比例如图 8.5 所示。

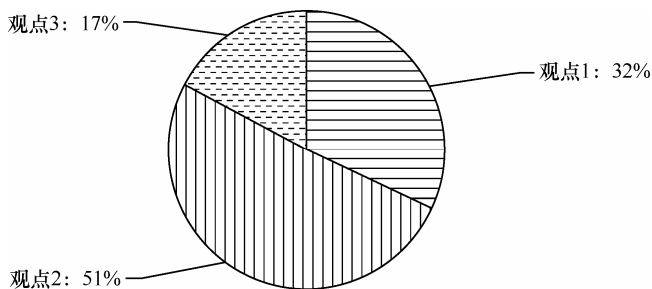


图 8.5 主题三细分子观点比例

#### 主题四：南海问题

在进一步的数据分析与研究中，对网民在这个主题内的深入观点进行梳理与分析，发现存在 3 个细分子观点。这些细分子观点所反映的内容如下：

观点 1. 71%的评论认为周边小的国家能敢于和中国对抗，肯定有大国的支持，不然他们这些国家考虑到经济实力、国土面积、国际地位等因素，也不会这么嚣张。

例如，用户“小品剧本创作者”评论道：“（菲律宾）全国军力不及中国南海舰队一半实力的菲律宾不这么认为！像强盗一样劫持了中国渔民，还要去审判！老鼠枕着猫腿睡，越来越大胆了！忠告一句，现代战争不是以前，一旦开火，一枚导弹就可以让一平方公里草木不生！当年俄罗斯的杜达耶夫怎么死的，也许有人一辈子忘不掉！话要说回来，既然中国是个大国，各种大杀器齐全，核弹、导弹、大驱逐舰、潜艇，清一色现代先进货，为什么没人拿中国像个大国样呢？时不时地挑衅，甚至调戏，困难时都来哭穷，困难过去立马不认识，越南甚至还对帮他活了命的中国大打出手！什么世道？文明的人类，为什么有这么多小人，忘恩负义的小人！到底是中国太善良了，还是地球上不要脸的太多了？？？有人说，中国是个养白眼狼专业户！我感觉，中国看错了这个世界！这个世界是弱肉强食，小国就是大国的衣食，菲律宾也眼馋日本，哭着喊着求美国兵去自己家里住，美国兵是省油的灯？节操呢？这个地球上的很多国家不喜欢善良型的国家，（国家）真正该善待的，是本国的公民！国际上的什么狗啊狼啊，忘恩负义的猴子啊，就得一手胡萝卜一手大棒子！和平相处可以有萝卜吃，调皮捣蛋就得打屁股！打就要让它记住教训！打败他就得让它求饶才行！”

观点 2. 20%的评论认为菲律宾支援朝鲜对抗中国只是虚张声势，考虑到自身的利益，也不会枉然出兵。

例如，用户“frankkee”评论说：“南海岛争，到底和谁开战最有利。爆发战争还有一个奇特的现象，就是频繁军演的区域就不会爆发真的战争，兵者诡道也，不可能在可能爆发实战的地方去军演，那样等于是告诉对方我准备何时、何地 and 用何种方法来打你，完全违背了战争发起的基本逻辑，所以在朝鲜半岛、伊朗和叙利亚周边持续不断的军演，反而一直斗而不破，没有发生真的战争。中国和周边国家的岛屿争端其实背后都有美国的因素，利用钓鱼岛，挑起中日对立，破坏东亚经济一体化，利用日本消耗中国，达到一石二鸟的目的，不会真的为了钓鱼岛和中国大打出手，所以钓鱼岛吵得热乎，其实是虚招，实的是

利用这个热点换取各方利益的最大化,大家没有必要再关注菲律宾而浪费太多精力。”

观点 3. 还有 9% 的评论比较激进,认为中国现在强大了,不是之前的中国,不要对这些小势力屈服,不惜动用武力。这种想法虽然比较激进,但是体现网民的爱国精神,更需要冷静理智分析。

例如,用户“致卓”评论道:“天涯友们,共同祈求解放军出兵南海还我尊严!”

综上,在本主题中,具有相应子观点的评论分布比例如图 8.6 所示。

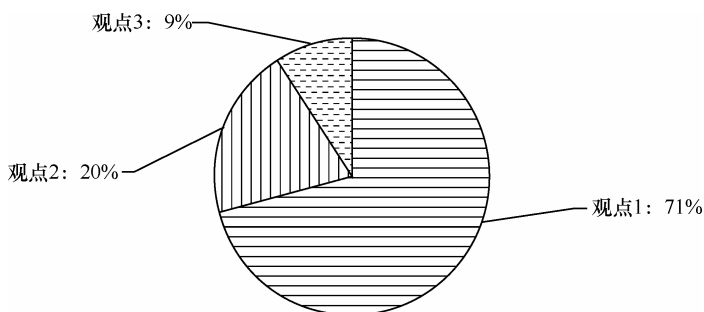


图 8.6 主题四细分子观点比例

### 主题五：伊拉克遗留的战争问题

在进一步的数据分析与研究中,对网民在这个主题内的深入观点进行梳理与分析,发现存在 3 个细分的子观点。这些细分的子观点所反映的内容如下:

观点 1. 38% 的评论伊拉克自从 2003 年被美国攻占以来,一直战乱不断,说明主权的重要性。

例如,用户“jszjjskjdx”评论道:“2003 年美国攻占伊拉克以来,一直战火不断,民间自发的保卫国家武装力量不断攻打在伊美国的政府部门,这是一个国家丧失主权后,国家的公民责任感的体现……”

观点 2. 30% 的评论认为 ISIS 就是美国养的一个“儿子”,美国不自己出面,让 ISIS 来搅乱国际治安。

例如,“龙王视野”评论道:“从某种程度来说,是这样的,美国为了打击叙利亚巴沙尔政权,曾帮助扶植过叙利亚周边的反政府势力,ISIS 是其中的一个,后来乘美国不注意发展壮大起来,这就是为什么与 ISIS 交火后的伊拉克士兵清理战场时找到了许多 M16 步枪,跟当年的本·拉登一模一样……”



观点3. 32%认为这是美国一种变相的侵略他国、霸占资源的手段,将来还会有其他的国家像伊拉克这样被以“正义”之名侵占。

例如,用户“战神之武士”评论道:“美帝在近几年的罪行,攻打伊拉克,搞乱叙利亚、乌克兰等一系列行径表示了美帝的野心,这只是个开端,将来会有很多国家会受美帝的无形侵占……”

综上,在本主题中,具有相应子观点的评论分布比例如图8.7所示。

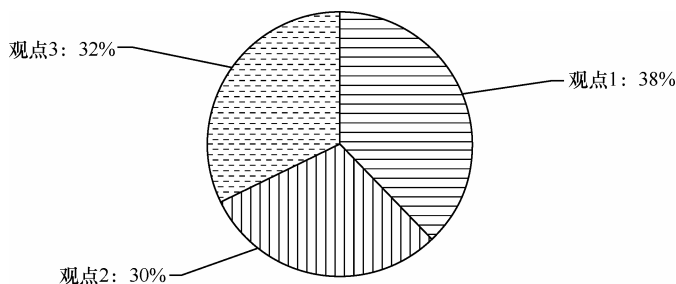


图 8.7 主题五细分子观点比例

### 3) 观点总结

领土与主权问题是每个人都关心的问题,领土与主权的完整是一个国家经济提升、国际地位提高,并获得飞速发展的基础,因此维护领土与主权的完整是全世界的渴望,作为一个普通公民,我们可以宣扬和倡导和平,积极维护党的号召,对于那些破坏国家领土与主权完整的行为,要不遗余力地指责和打击。只有这样,世界才会真正的和平,人民才会过上幸福的生活。

#### 8.3.3 专题传播趋势

传播性是网络舆情的一个非常鲜明的特性。本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量,专题情感分析的基础是帖子回复内容的情感趋向,涉及情感计算。

##### 1) 影响力

我们认为流量与影响力成正比,流量越大影响力越大。为此,本节分别对新增回复数对应每天的变化和总回复数对应每天的变化做可视化得到波线图(图8.8,图8.9),以便发现对应的发展峰值。

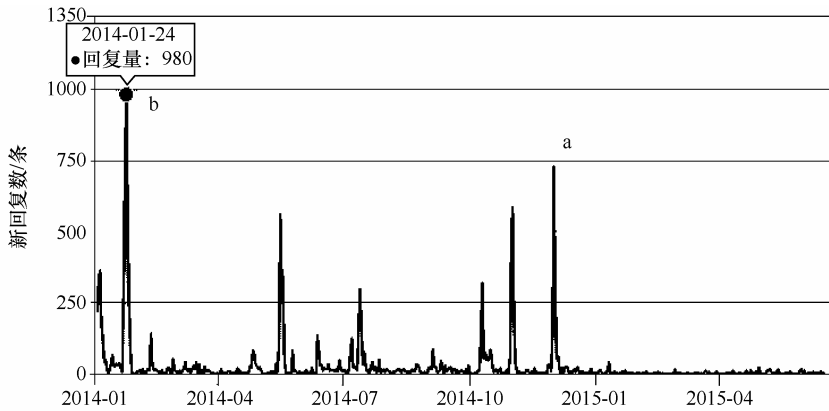


图 8.8 专题日流量时间趋势

点 a：2014 年 1 月 24 日；点 b：2014 年 12 月 1 日

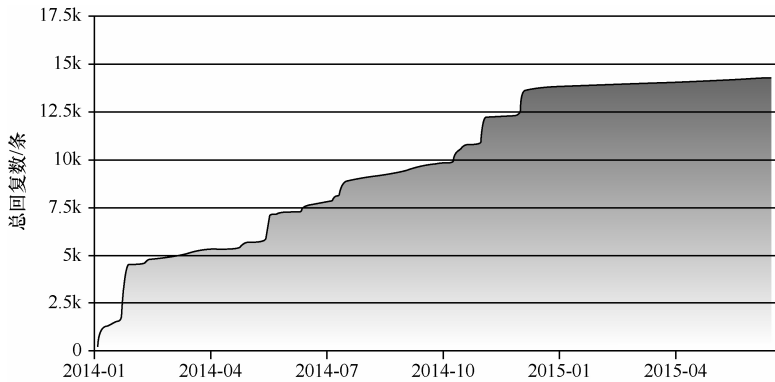


图 8.9 专题总流量时间趋势图

从图 8.8 可以看出每天对应的新增回复数的变化。在点 b 之前，每天对应的新增回复量，有时增加，有时下降，但相对来说都是比较低并且稳定，说明这一段时间该事件提出来之后并没有引发很大的影响；而在 1 月 20 日之后，每日新增量开始瞬间增高，并持续持续上涨趋势，在点 b 达到了发展峰值，说明该专题在 1 月 20 日后引发的讨论越来越激烈，在点 b 到达了最激烈的点。在点 b 之后引发的讨论稍微下降，但是也有上涨趋势，在点 a 又达到一个峰值，说明这事件在这个时间又被提起，说明很可能是一个在 a 点对应的日期所发生的事件引起了 b 点对应事件发生的事件的再次讨论，进一步说明，这两个时间点对应的事件是相似事件。

2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库，对专题内的回复内容进

行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图（图 8.10）和专题情感分类图（图 8.11）。

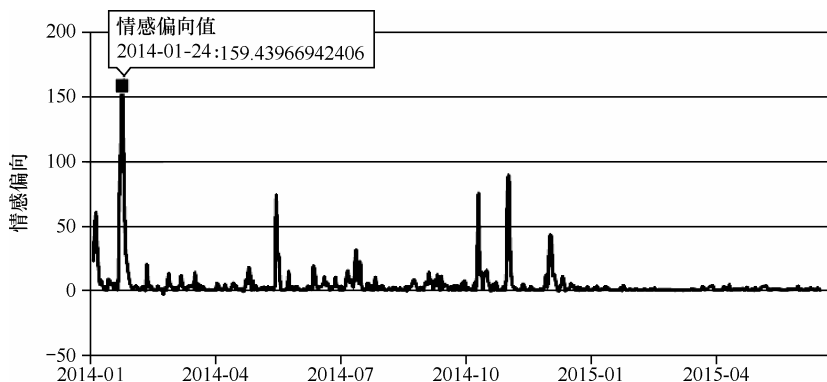


图 8.10 专题情感趋势图

点 a: 2014 年 1 月 20 日; 点 b: 2014 年 1 月 24 日

从图 8.10 可以看出发展峰值变化情况。例如，在点 a 对应的日期所对应的情感偏向的情感值比较低，说明这段时间的情感波动不明显；到达点 a 时候，情感急速上升，在点 b 达到峰值，说明这些点属于情感拐点，必定有一些事情影响了网民的情感。除此之外，还可以看出在领土与主权领域网民的情感极易受到影响。结合图 8.8 和图 8.10 可以发现一些规律：①图 8.10，1 月 24 日达到情感发展峰值，图 8.8 中 1 月 24 日之前日增长量比较少，说明该日之前的情感大多数是正值。②图 8.8 中的影响力发展峰值对应的日期情感值也同时达到峰值，说明流量和情感值也有一定的正相关性。

从图 8.11 可以看出，在网民的评论中，负面评论占的比重较大，说明爱国主义者听到有损国家主权的事件后，在网上掀起了一波爱国热潮。

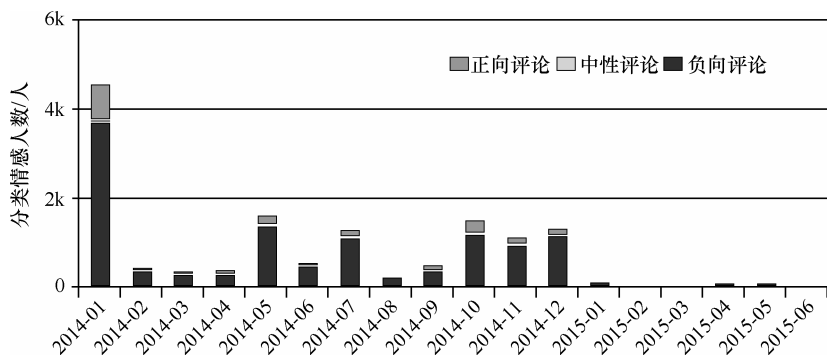


图 8.11 专题情感分类图

### 8.3.4 专题用户情感的分析

探索网民情感可以发现情感积极的用户，可以找出产生消极影响的网民，对网络舆情监控具有重要意义。

本节对各专题下的人物情感进行了统计分析，由于活跃度对情感的影响较大，所以本节分别以情感值和活跃度为两个坐标轴构建专题任务散点图（图 8.12）。

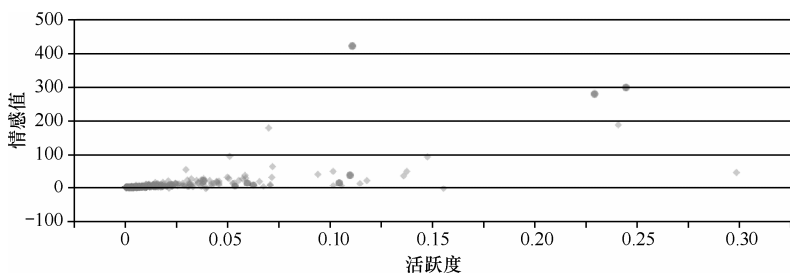


图 8.12 基于活跃度和情感值的用户散点图

圆点代表发帖人，方点代表回帖人

从图 8.12 中可以看出，一般发帖人的情感值会较高一些，另一些点不是发帖者但其活跃度较高。这两种可能，一种是该用户作为评论人，评论的内容又引发了一个新的讨论热潮，另一种可能就是属于那些发帖较少，经常参与那些评论或者原帖回复的那一类人，这些人有一定的回复参与量，但是通常回复的情感极性较低，感情色彩不明显，这些人所属领域性比较明显。

为了进一步进行分析，本节将那些情感值负性较大且活跃度较小的人进行了详细作图描绘，取情感值小于 10，活跃度大于 0.02 的所有用户，作图 8.13。这类用户代表的是那些在论坛活动积极，但是负面情绪较大的那些人。由于这些人活跃度相对较大，所以他们影响其他用户的可能性也较大，会给论坛用户灌输一些负面的情绪。因此，从这类人群中容易发现一些用户，并对他们进行权限的限制、屏蔽、删除。

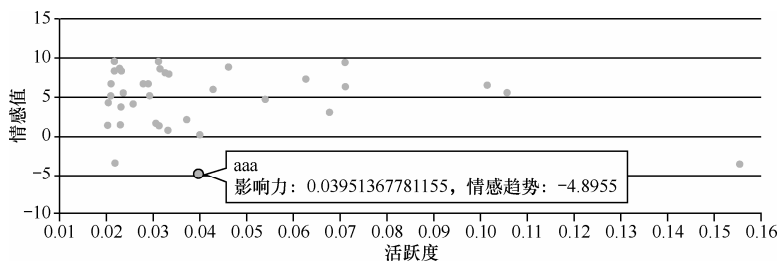


图 8.13 低情感和高活跃度散点图

图 8.13 所标出的点可能是那类宣扬消极情绪的用户，论坛管理员要对这些用户进行监察和历史查询，以便进行一系列的操作。

8.3.5 专题网络的构建与分析

构建专题网络的方法见节 4.3.5。构建领土与主权的专题网络，进行可视化的结果见图 8.14。

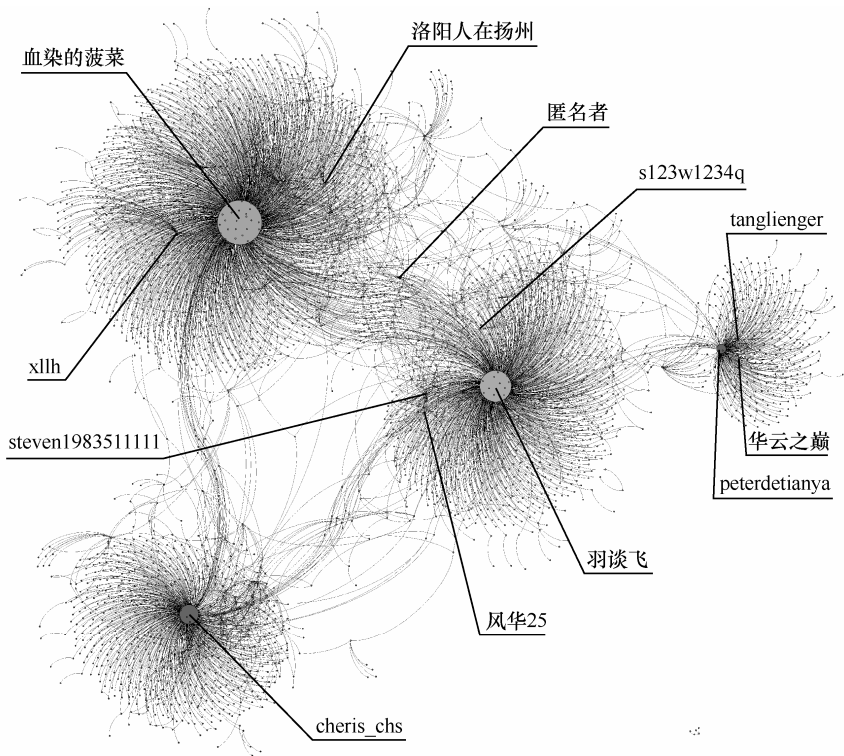


图 8.14 专题传播网络

1) 网络指标概述

本节分别从社会网络的整体性指标（表 8.8）、中心性与影响力指标、关联系与“小世界”指标进行了分析，以便发现网络中的特殊节点，以便分析个体网络演化与传播。

表 8.8 网络基本指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
2.614	0.001	7	0.701	3.384	0.591

平均度：较小，平均加权重为 2.614；图密度：该传播网络图比较稀疏；网络直径：较长，该网络的传播较远；模块化指标：划分出 7 个社区，且划分的质量较高；平均路径长度：较大，网络传播较广；平均聚类系数：聚类效果不错。

2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算, 这里取点度中心度排名前 6 的节点, 列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名, 得到对应的数据如表 8.9 所示。

表 8.9 中心度指标排名

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	值排名	绝对值	相对值	值排名	绝对值	相对值	值排名
血染的菠菜	2125	0.225	1	1349704	12.081	1	37822	8.839	2
羽谈飞	1722	0.182	2	1224862	10.963	2	38163	8.760	3
cherish_chs	1059	0.112	3	766938	6.865	3	39300	8.506	8
peterdetianya	1052	0.111	4	443251	3.967	4	42437	7.878	243
tanlienger	325	0.034	5	740	0.007	183	45762	7.305	458
华云之巅	250	0.026	6	7415	0.066	96	45758	7.306	456

从表 8.9 中可以看出“血染的菠菜”“羽谈飞”等 6 个节点的点度中心度排名前 6 位。从中间中心性来看, “血染的菠菜”的中间中心度远远大于其他节点, 排在第一位, 说明许多节点都需要经过该节点才能建立联系并获取信息。从接近中心性来看, “血染的菠菜”和“羽谈飞”的接近中心度排名前两位, 说明这两个节点与其他节点进行交互比较容易。

图 8.14 的点度中心势指标为 0.22%, 该指标值很小, 表示该图在度数上向某点集中的趋势很不明显; 该图的中间中心势指标为 12.06%, 该指标值也比较小, 表示该图向某些中间节点聚集的程度比较小, 即在整个网络中大部分的节点并不需要经过某些中间节点, 就能得到信息。

3) 关联性与“小世界”

图 8.14 的特征途径长度为 4.258, 凝聚度为 0.058。该图路径长度的具体分布如表 8.10 所示。

表 8.10 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	4879	0.002	6	365358	0.151
2	321055	0.132	7	84067	0.035
3	280538	0.116	8	10984	0.005
4	826268	0.340	9	333	0.000
5	533823	0.220			

从表 8.10 可以看出,可见路径长度为 4 占的比例最大,并且路径长度介于 4~6 的比例占了一半以上,说明该网络下大部分路径长度都在 4~6,分布较为分散。

图 8.14 的聚类系数为 6.966,加权的聚类系数为 0.010。节点聚类系数如 8.11 表所示。

表 8.11 节点聚类系数

用户	聚类系数	点对数
血染的菠菜	0.000	979300
羽谈飞	0.001	476776
cherish_chs	0.002	167910
peterdetianya	0.002	33411
xllh	0.023	1035
洛阳人在扬州	0.366	325

8.3.6 个体网络分析

按照整体网的分析之后,选择一些传播特性相对较强的用户做一个帖子下所有用户的个体传播网络。本节选取“羽谈飞”所发的帖子下对应的所有用户,对它进行指标分析,并且描述传播路线。图 8.15 为通过用户发表的帖子中用户的评论关系构建的传播网络。

1) 结构洞

这里列举节点度数排名前 5 位的用户的结构洞指标,形成相应的数据指标表(表 8.12)。

表 8.12 结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
羽谈飞	978	926.068	0.947	0.226	0.956
风华 25	49	37.507	0.765	0.399	0.659
steven1983511111	14	7.868	0.562	0.737	0.594
匿名者	13	8.739	0.672	0.712	0.700
s1234w1234q	13	5.591	0.430	0.882	0.840

从“有效规模”列看出“羽谈飞”所对应的值最大,达到了 926.068,这说明“羽谈飞”最居于网络的核心。“效率”列对应的值最高的为“羽谈飞”,达到了 0.947,说明该用户在社会网络中交际最广,如果散发消息的话,理论上应该是最快的。“限制度”列看出,“s1234w1234q”对应的值最大,为 0.882 说明该用户在网络中受到的限制最大。

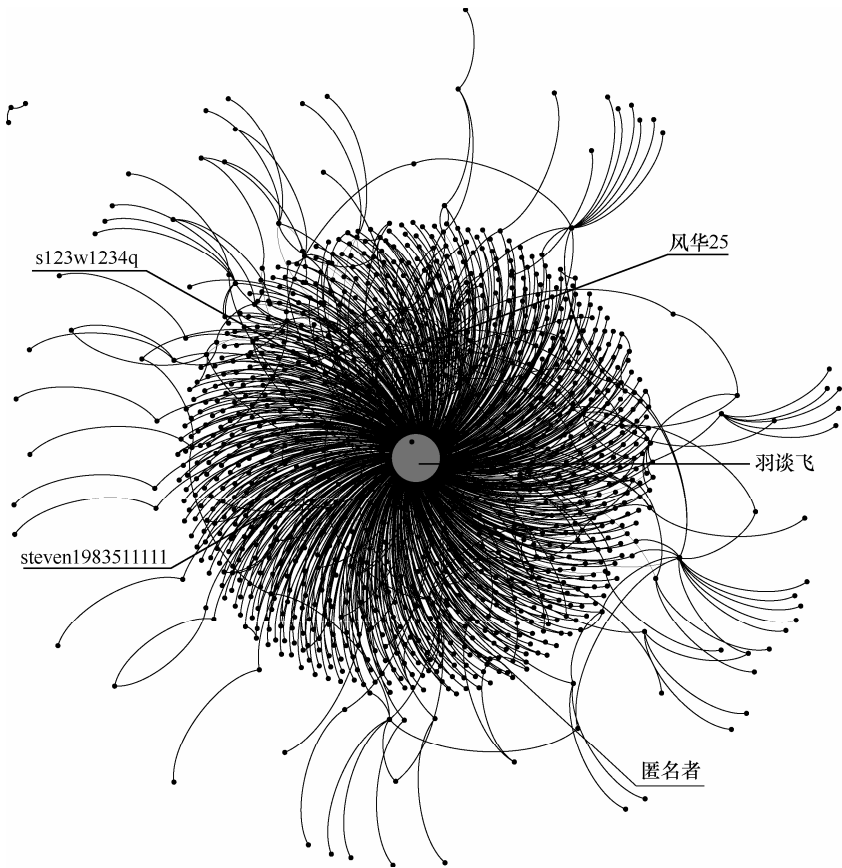


图 8.15 个体网传播图

2) 诚实中间人

从表 8.13 中可以看出,“羽谈飞”代表的节点作为中介的次数为 116 次,共联络了 270 对关系,其中 235 对关系属于 HBIO 类,23 对关系属于 HBIO1 类,12 对关系属于 HBIO2 类。这说明大部分节点属于纯中介,说明参与回复主帖的人物之间的关系比较弱。同时也表明大部分回复是针对主帖回复的,而只有一小部分回复针对的是评论用户评论的。

表 8.13 诚实中间人

用户	Size	Pairs	HBIO	HBIO1	HBIO2
羽谈飞	116	270	235	23	12
风华 25	18	153	126	17	10
s1234w1234q	5	10	7	2	1

HBIO 代表纯中介, HBIO1 代表弱中介, HBIO2 代表非中介。



从图 8.15 中可以看出,“羽谈飞”发表的帖子后,大量用户参与的讨论;在回复的用户中,“风华 25”与“s1234w1234q”发表的评论引起了网民大量回复,说明该类用户在网络传播中可能是一个意见领袖,他的主观意见会影响或者左右其他用户的情感倾向。

8.4 专题 2：合作共赢

从古老的石器时代,我们的祖先为了捕捉猎物,往往趋向于几个人之间的合作,来获取足够的食物。在动物世界中,单个狼的力量不足为患,但是当它们聚集在一起组成一群狼的时候,恐怕是动物中最恐怖的一件事情。由此可见,合作的目的是为了获得更好的利润,达到大家获益的目的。当今社会,一些弱小的发展中国家为了提高自己的国际地位,往往多个国家组成一个大的体系,成为一股世界上大国都不敢忽视的力量。

本专题探究合作共赢,今年召开的 APEC 峰会、G7 峰会等一系列的国家之间合作大会,说明了合作的重要性,本专题的探索具有十分重要的意义。

8.4.1 专题发展简介

在经济飞速发展发当今社会,国家之间不像以前那样,与外界隔绝,孤立式发展,或者很少和外界协作,这种方式被证明是错误的发展方式。进入 21 世纪,各国为了提高国家的国际地位,增强国际竞争力,促进国家经济发展,往往积极寻求和外界交流机会,因此一些联盟组织和一些经济共同体出现,如“欧盟”“北约”“亚太”等一系列组织,又如 2014 年举行的“APEC 峰会”“金砖五国会议”“非盟首脑会议”等合作性会议。这些组织形成或者会议的召开都是为了促进国家的经济合作发展,提高国际地位,建立合作伙伴。

表 8.14 是合作共赢主题下的几大主要热帖。在帖子的专题传播过程中,大量的网民针对所发表帖子的观点进行了回复,其中也产生了一些新的且具有一定影响力的观点,并引发了网民的关注和讨论;尽管新观点基本还是发帖者的观点延伸,但也体现了网友们讨论的内聚性和聚焦性。

表 8.14 合作共赢专题热帖

帖子标题	发帖人	发帖时间	点击量/ 次	正面 回复 量/条	中性 回复 量/条	负面 回复 量/条	作者 情感
忠言逆耳,讲点中国人不爱听的话吧,这次中国人访问欧洲效果几乎为零!	龙王 视野	2014/04/03	93809	251	361	64	100.000

续表

帖子标题	发帖人	发帖时间	点击量/ 次	正面 回复量/条	中性 回复量/条	负面 回复量/条	作者 情感
普京一箭三雕，西方输惨了（转载）	雨痕疏影	2014/09/03	10851	35	59	8	87.382
有没朋友看到昨晚 APEC 晚宴上，中.俄.美.日领导人互动的精彩之处？	诸葛隆中	2014/11/11	590762	492	579	141	82.523
金砖银行落户上海，敲响了美元霸权的丧钟！！	刘育新 2013	2014/07/19	30673	186	133	37	66.648
旗扬时事分析	旗扬 2014	2014/11/14	1116951	514	633	87	9.4491

8.4.2 观点分类与观点列表

通过上述的网络数据信息，针对网民在回复过程中的观点挖掘与倾向进行语义的处理，形成了本话题内的核心观点。

1) 观点分类

分析合作共荣专题中网民在帖子中发表的全部言论，经过文本清洗、语义分析、观点聚类步骤，得到的三个核心主题。各主题观点所占比例如图 8.16 所示。

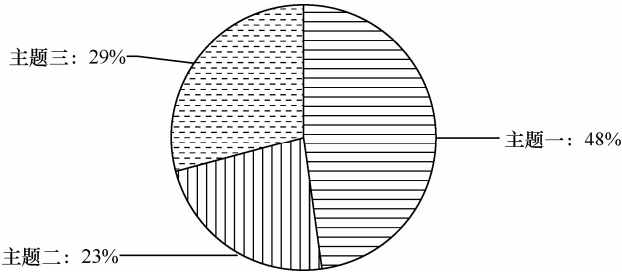


图 8.16 “合作共荣”专题观点分类比例

- 主题一：北京召开 APEC 峰会，美国和西方国家散播中国威胁论
- 观点 1. 中国人口那么多，中国人富裕了，会威胁到美国的地位。
- 观点 2. 美国在乎的不是中国经济超过美国，而是怕中国经济超过美国后，

会威胁到美国的霸主地位。

观点 3. 美国其实之前是促进中国经济的, 不然早就对中国进行经济打压了。

### 主题二: 金砖五国峰会上提出建设金砖银行

观点 4. 这是国际版的农村包围城市的路线。

观点 5. 摆脱了美元对发展中国家的限制, 人民币的地位将会大大提高, 成为能和美元和欧元抗衡的货币。

观点 6. 金砖银行总部设在了上海, 上海很有可能超越新加坡、中国香港等成为亚洲的金融中心, 甚至将来会成为纽约级别的世界经济中心。

### 主题三: G7 峰会上, 点名中国的海域问题并指责中国

观点 7. 清者自清, 不用做过多的解释。

观点 8. 中国的海域问题都是美帝唆使的走狗国家的叫嚣行为, 目的打击中国。

观点 9. 中国在南海问题上不能认输, 这是中国一个态度的表现。

## 2) 观点分类详细分析

为了更好地对相应的观点分类进行深入的分析, 本节对每个分类观点内的网络用户分布以及网络民意进行了分类处理, 通过网络用户真实反馈的信息来反映他们对这一些问题的思考与建议。

### 主题一: 北京召开 APEC 峰会, 美国和西方国家散播中国威胁论

观点 1. 41%, 中国人口那么多, 中国人富裕了, 会威胁到美国的地位。

试想一下, 如果中国人人富裕, 人均 GDP 超过美国, 加上中国的人口是美国的四倍多, 那么难免会影响到美国的地位。所以, 美国会散播一些中国威胁论, 来恐吓和威胁与中国合作的国家以切断中国与合作伙伴的联系。

例如, 用户“至尊郭嘉\_”的评论说: “废话, 中国人这么多, 都过上好日子, 你让美国人怎么过”。

观点 2. 25%, 美国在乎的不是中国经济超过美国, 而是怕中国经济超过美国后, 会威胁到美国的霸主地位。

我们知道, 美国是一个经济大国, 当然也是军事的霸主。中国作为一个改革开放取得重大成就的国家, 经济上升速度飞快, 当中国经济增长到超过美国经济的时候, 美国就会认为那个时候中国就会成为世界的霸主。可是, 中国奉行的政策是和平外交政策, 即使到了那一步, 中国不会称霸。

例如, 用户“王佐 2013”评论说: “不是, 美国不是在乎你经济上超越美国, 而是在乎你经济上超越了美国以后要采取军事霸权, 要开始和美帝争夺世界霸权”。

又如, 用户“hechengwang”评论说: “你要看深层次原因, 中国人富了不要紧, 只要你不抢我的全球霸主地位就行了, 权利在我手上就行了, 就像日本人一样做狗, 当狗使唤, 当奴隶压迫”。

观点 3. 34%, 美国其实之前是促进中国经济的, 不然早就对中国进行经济打压了。

这类用户持有的观点是, 如果美国要是想打压中国的经济, 那么就早对中国的经济进行打压了。所以说, 美国一开始是促进中国经济的。对于美国, 中国的劳动力市场和资源是美国合作所必需的; 当中国经济达到一定程度后, 有某些方面威胁到美国地位的时候, 美国才开始实行压制策略。

例如, 用户“潇洒小狼”评论说: “要是美国不愿意给你中国发展经济的机会? 美国早把你打压了。”

综上, 在本主题中, 具有相应子观点的评论分布比例如图 8.17 所示。

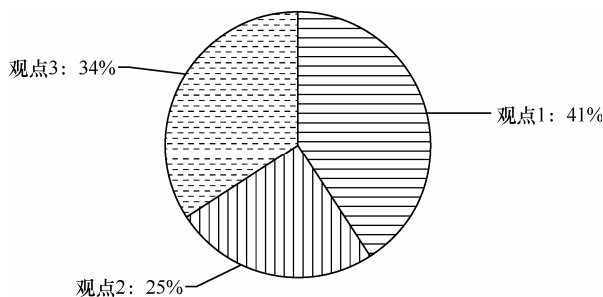


图 8.17 主题一细分子观点比例

## 主题二：金砖五国峰会上提出建设金砖银行

观点 1. 21%, 这是国际版的农村包围城市的路线。

持有这类观点人将这一形式形象地比喻成共产党抗战时期的农村包围城市的路线, 将美国比喻成富有的国民党, 将这些发展中国家比喻成共产党。

例如, 用户“174947392”评论说: “新形势下的联合政府, 国际版的农村包围城市。”

观点 2. 33%, 摆脱了美元对发展中国家的限制, 人民币的地位将会大大提高, 成为能和美元和欧元抗衡的货币。

美元被公认为最流通货币, 那是因为美国处于霸权地位。金砖银行的建立, 人民币会在金砖国家内流行, 加速了人民币的流通性, 也成为反美国霸权的一个筹码。

例如,用户“只议不论”说:“这个好,摆脱美元的控制。下一步四国本国货币结算。”

观点3. 46%,金砖银行总部设在了上海,上海很有可能超越新加坡、中国香港等成为亚洲的金融中心,甚至将来会成为纽约级别的世界经济中心。

银行象征的就是经济中心性,金砖银行设立在上海,说明将来上海的经济地位会飞速提升,加速了中国经济的步伐。

例如,用户“铁幕与乌云 ZH”说:“国家设立上海自由贸易区,现在金砖银行总部又可能设在上海,可以肯定地说,只要中国与金砖国家的经济进一步发展,这对中国与上海来说是极大的利好。”

上海在将来可能会超过新加坡、中国香港等成为亚洲金融中心,甚至不排除随着中国经济规模占世界经济比重越来越大、人民币完全开放成为世界货币,那时候上海对世界的金融影响有追赶伦敦、纽约的趋势。

可以肯定的是,这对一直把金融地位看得很重的中国香港来说,绝不是什么好消息。如果上海的金融地位得到进一步开放与加强,中国香港势必会被边缘化,最好的结果就是,如果把上海比作纽约的话,那中国香港只能是芝加哥。”

综上,在本主题中,具有相应子观点的评论分布比例如图 8.18 所示。

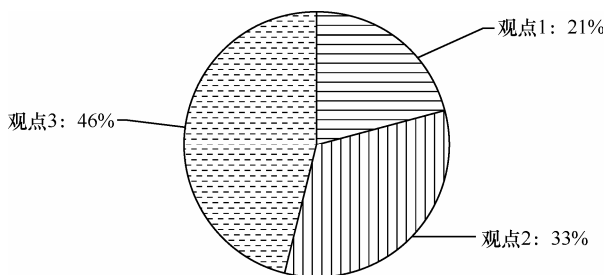


图 8.18 主题二细分子观点比例

### 主题三：G7 峰会上，点名中国的海域问题并指责中国

观点1. 19%，清者自清，不用做过多的解释。

对于南海问题，各方都有自己的说辞，只要坚信自己是正确的，中国就不应该理会这件事情，坚持自己国家的主权完整。

例如，用户“未念念”评论说：“不卑不亢、不称兄道弟、不称霸但不手软。走好自己的路，不在乎人家怎么说。”

观点2. 48%，中国的海域问题都是美帝唆使的走狗国家的叫嚣行为，目的打击中国。

中国的海域问题不是因为中国强大了而实行的一种霸占海域的行为。南海问题签署协议时候,本来就是中国的领土。中国有权利做事情,有足够的理由来容纳它,而周边的国家没有那种胆量,在自己没有理的情况下还来争夺、军事演习等,一定是背后有大国支持他们,他们才有恃无恐地进行叫嚣。

例如,用户“爱国者”评论说:“G7 会议是强国的游戏,是划分地球,霸占资源的会议,当然他们会提出一些对自己有利的建议和问题,有七个经济强国做后台,一起施加压力;但是中国现在也不是弱国,凭一些小的国家是不敢举行军事演习之类的,肯定是走狗国,谢谢。”

观点 3. 33%,中国在南海问题上不能认输,这是中国一个态度的表现。

中国南海问题只是一个开端,如果说南海问题中国妥协了的话,那么中国周边的其他问题也会接踵而来,这是中国态度的一个表现。中国态度一定要强硬,才不能让其他国家抓住把柄。

例如,用户“ALM375”评论说:“中国南海问题只是一个开端,如果说南海问题中国妥协了的话,那么中国周边的其他问题也会接踵而来,这是中国态度的一个表现。中国态度一定要强硬,才不能让其他国家抓住把柄。”

综上,在本主题中,具有相应子观点的评论分布比例如图 8.19 所示。

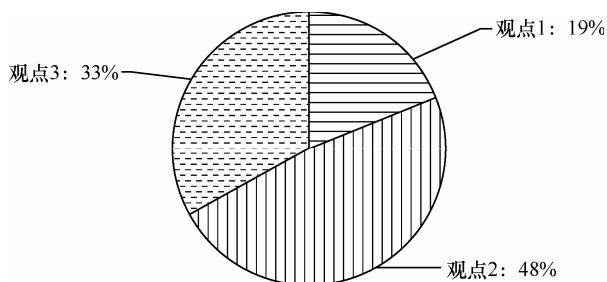


图 8.19 主题三细分子观点比例

### 3) 观点总结

从本专题可以看出,关注 APEC 峰会的人数比例比较大,网民的情感往往比较正向,正面情绪比较大;而关注金砖五国的会议的人数相对来说较少一些,用户可能对此认知不够;而对于 G7 峰会上点名指责中国的南海问题这一事件,讨论比较广泛,说明我国大多数网民具有爱国主义情怀和强烈的民族责任感,当他们了解这一个事件后,在网络上发泄自己的爱国情怀,引起了广泛的讨论。

### 8.4.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

#### 1) 影响力

本节分别对新增回复数和总回复数按照时间进行统计，然后对统计结果进行可视化得到波线图（图 8.20，图 8.21），以便发现专题传播的发展峰值。

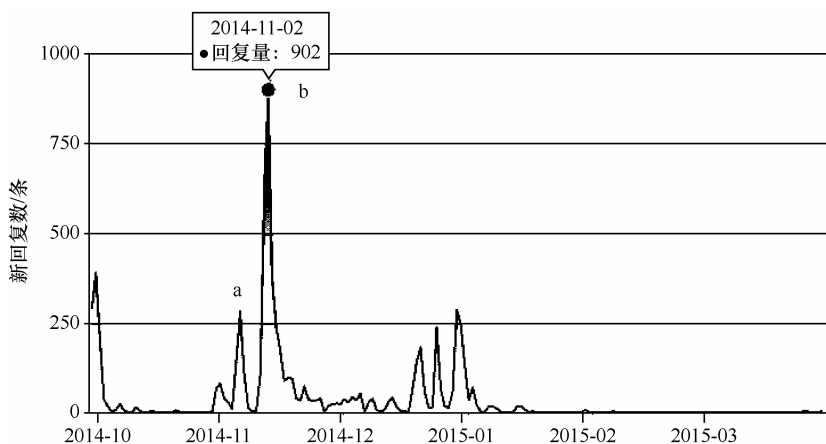


图 8.20 专题日流量时间趋势图

点 a: 2014 年 11 月 5 日; 点 b: 2014 年 11 月 2 日

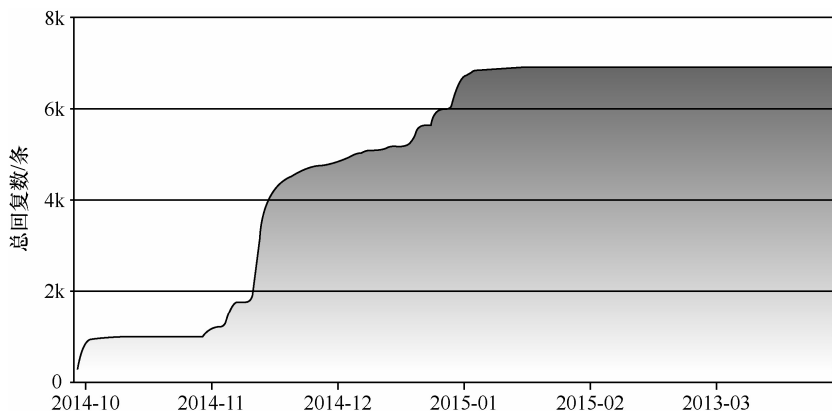


图 8.21 专题总流量时间趋势图

从图 8.20 可以看出每天对应的新增回复数的变化情况。在点 a 之前, 每天对应的新增回复量, 有时增加, 有时下降, 但相对来说都是比较稳定并且相对来说比较稳定, 说明这一段时间该事件提出来之后并没有引发很大的影响; 而在点 a 之后, 每日新增量开始瞬间增高, 并持续持续上涨趋势, 在点 b 达到了发展峰值, 说明该专题在点 a 后引发的讨论越来越激烈, 在点 b 到达了最激烈的点。在点 b 之后虽然评论总量有所增加, 但是日增长率急剧下降。对于峰值 b 点, 我们猜想可能是在该天发生了一个事件, 而引发了人们激烈的讨论。点 b 对应的时间过后, 说明该事件慢慢淡出了人们的视线。但是不保证将来不会被其他事件牵连而再次被提起。

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库, 对专题内的回复内容进行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图 (图 8.22) 和专题情感分类图 (图 8.23)。

从图 8.22 可以看出发展峰值和发展拐点。点 a 和点 b 对应的情感偏向的情感值比较高, 说明这两个时间点达到了情感的发展峰值; 而在点 a 之后, 情感开始下降, 经过一段时间后, 开始上升, 到达点 c 的峰值; 然后接着下降, 直到点 b 达到最大值; 点 b 之后, 就下降了, 随后达到了平稳, 说明这些点属于情感拐点, 必定有一些事情影响了网民的情感。除此之外, 还可以看出在合作共赢专题内的网民的情感极易受到影响。

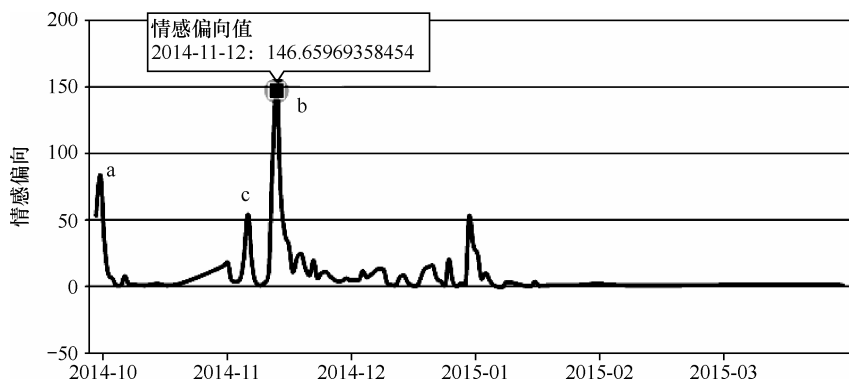


图 8.22 专题情感趋势图

点 a: 2014 年 9 月 30 日; 点 b: 2014 年 11 月 5 日; 点 c: 2014 年 11 月 12 日



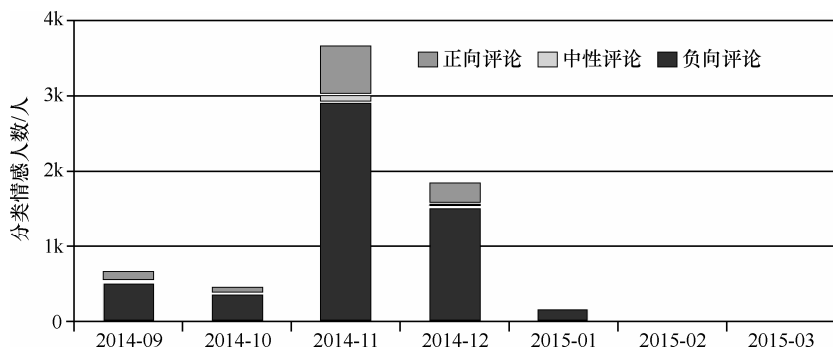


图 8.23 专题情感分类图

从图 8.23 可以看出，在网民的评论中，负面评论占的比重较大，说明现在网络环境下，喷子较多，网民发言还是比较随意，大部分网民选择在网络上发泄自己的情绪。

#### 8.4.4 专题用户情感的分析

探索网民情感可以发现情感积极的用户，可以找出产生消极影响的网民，对网络舆情监控具有重要意义。

本节对各专题下的人物情感进行了统计分析，由于活跃度对情感的影响较大，所以本节分别以情感值和活跃度为两个坐标轴构建专题任务散点图（图 8.24）。

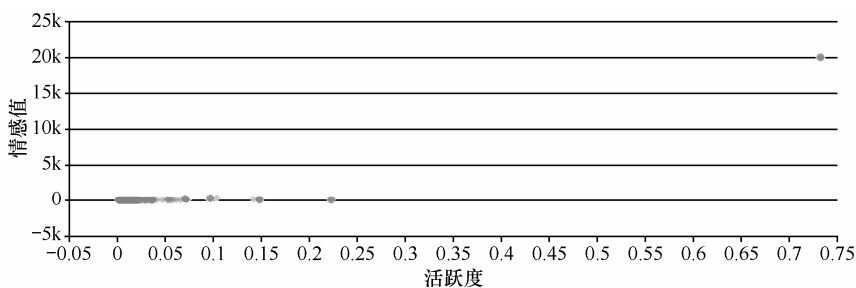


图 8.24 基于活跃度和情感值的用户散点图

圆点代表发帖人，方点代表回帖人

从图 8.24 中可以看出，一般发帖人的情感值都比较高一些，而那些红点的情感值比较高且活跃度也很大，所以这些用户很可能是意见领袖。但是也存在一些点不是发帖者但是其活跃度较高，说明有两种可能，一种是该用户作为评论人，评论的内容又引发了一个新的讨论热潮，第二种可能就是属于那些发帖较少，经常参与那些评论或者原帖回复的那一类人，这些人有一

定的回复参与量，但是通常回复的情感极性较低，感情色彩不明显，但是这些人所属领域性比较明显。

为了进一步进行分析，本节将那些情感值负性交大且活跃度较小的人进行了详细点的描绘，取情感值小于 20，活跃度大于 0.02 的所有用户，作图 8.25。这类用户代表的是那些在论坛活动积极，但是负面情绪较大的那些人，这些人的发言或者参与会影响到人们对于事件情感的判断。由于他们的活跃度相对来说较大，所以他们影响其他用户的可能性比较大，会给论坛用户灌输一些负面的情绪。因此，很容易从这类人群中发现一些用户，并且对其进行权限的限制、屏蔽、删除。

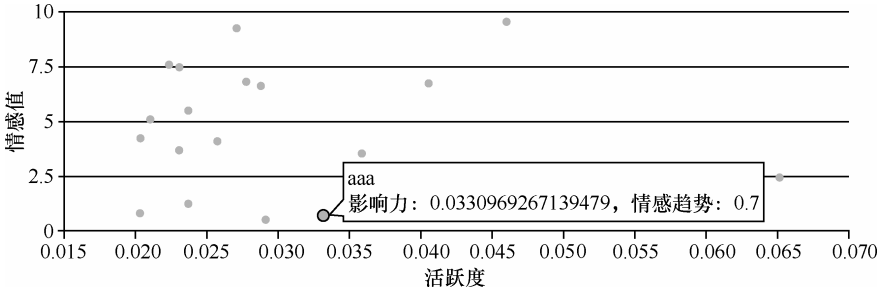


图 8.25 低情感和高活跃度散点图

图 8.25 有些点比较特殊，如“aaa”，他们都是相对活跃度比较大、但是情感值比较低，这类用户极有可能是负面情绪的发起者，论坛管理员应该采取相应的操作。像活跃度为 0.033、但其情感值却只有 0.7 的用户，极易被利用，发展成为消极情绪的发起者，应予以关注。

### 8.4.5 专题网络的构建与分析

构建专题网络的方法见节 4.3.5。构建合作共荣的专题网络，进行可视化结果见图 8.26。

由图 8.26 可知，“诸葛隆中”和“邗一”发布的帖子及受众构成了该主题下的两个最大的社区。此外还有一些用户发表的帖子形成了一些小的社区。在以“诸葛隆中”为中心点的社区中，可以在图的左上角中看到最典型的一条传播线，“你还我钱”的用户对原帖进行了评论，然后“整整齐齐大圣”对其进行了回复，引起了另外一群人的关注的讨论。而在以“旗扬 2014”为中心点的社区中，则较少有其他度数较大的节点，这说明没有出现新的话题，从而引起一部分人讨论。

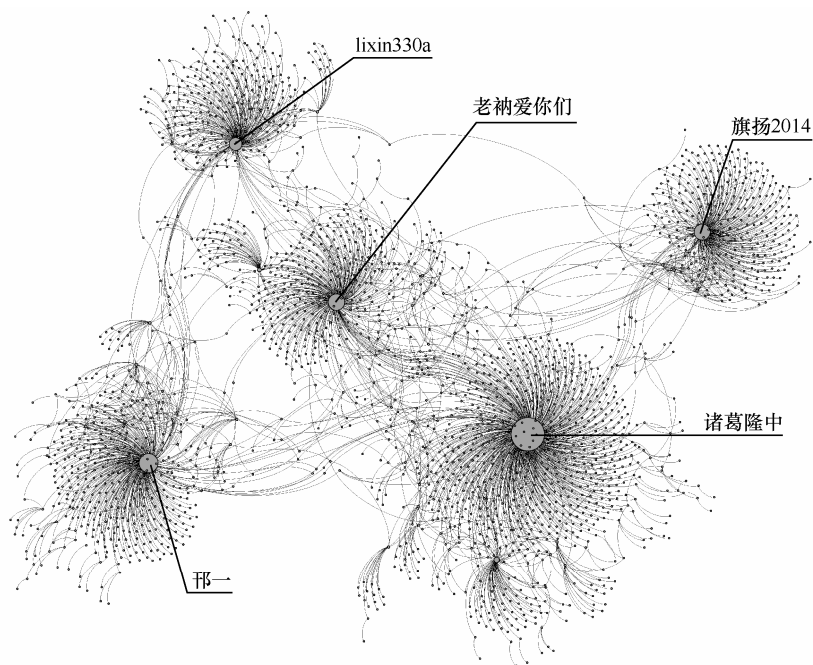


图 8.26 专题传播网络

1) 网络指标概述

本节分别从社会网络的整体性指标（表 8.15）、中心性与影响力指标、关联系与“小世界”指标进行了分析，以便发现网络中的特殊节点，以便分析个体网络演化与传播。

表 8.15 网络基本指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
2.577	0.001	9	0.764	3.775	0.447

平均度：较小，平均加权重为 2.577；图密度：该传播网络图比较稀疏；网络直径：较长，该网络的传播较远；模块化指标：划分出 7 个社区，且划分的质量较高；平均路径长度：较大，网络传播较广；平均聚类系数：聚类效果不错。

2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算，这里取点度中心度排名前 5 的节点，列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名，得到对应的数据如表 8.16 所示。

表 8.16 中心度指标排名

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	排名	绝对值	相对值	排名	绝对值	相对值	排名
旗扬 2014	915	0.564	1	274871	5.422	7	59466	3.787	161
诸葛隆中	798	0.492	2	362755	7.156	4	50164	4.489	5
邗一	675	0.416	3	472950	9.330	1	50159	4.490	4
lixin330a	461	0.284	4	359256	7.087	5	55329	4.070	61
老衲爱你们	363	0.224	5	124819	2.462	9	50934	4.421	6

从表 8.16 中可以看出“旗扬 2014”“诸葛隆中”等 5 个用户的点度中心度排名前 5 位。从中间中心性来看,“邗一”的中间中心度远远大于其他节点,排在第一位,说明许多节点都需要经过该节点才能建立联系并获取信息。从接近中心性来看,“邗一”和“诸葛隆中”的接近中心度排名前两位,说明这两个节点在网络中很可能做“桥”的作用,这些点代表的人一般具有人缘好、交际广等特点,同时这些用户很容易被人利用来传播垃圾信息。

图 8.26 的点度中心势指标为 0.56%,该指标值很小,表示该图在度数上向某点集中的趋势很不明显;该图的中间中心势指标为 9.30%,该指标值也比较小,表示该图向某些中间节点聚集的程度比较小,即在整个网络中大部分的结点并不需要经过某些中间节点,就能得到信息。

### 3) 关联性与“小世界”

图 8.26 的特征途径长度为 7.116,凝聚度为 0.024。该图路径长度的具体分布如表 8.17 所示。

表 8.17 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	3109	0.005	10	73444	0.113
2	50269	0.077	11	42681	0.066
3	32938	0.051	12	13889	0.021
4	46408	0.071	13	22439	0.034
5	82278	0.126	14	12731	0.020
6	74049	0.114	15	3524	0.005
7	90227	0.139	16	478	0.001
8	66142	0.102	17	28	0.000
9	35931	0.055			

图 8.26 的聚类系数为 1.301,加权的聚类系数为 0.004。节点聚类系数如 8.18 表所示。

表 8.18 节点聚类系数

用户	聚类系数	点对数
诸葛隆中	0.000	218130
旗扬 2014	0.002	45150
邗一	0.002	65341
lixin330a	0.002	27261
老衲爱你们	0.000	47586

8.4.6 个体网络分析

按照整体网的分析之后，选择一些传播特性相对较强的用户做一个帖子下所有用户的个体传播网络。本节选取“诸葛隆中”发表的一篇帖子，对它进行指标的分析，并且描述传播路线。图 8.27 为通过用户发表的帖子中用户的评论关系构建的传播网络。

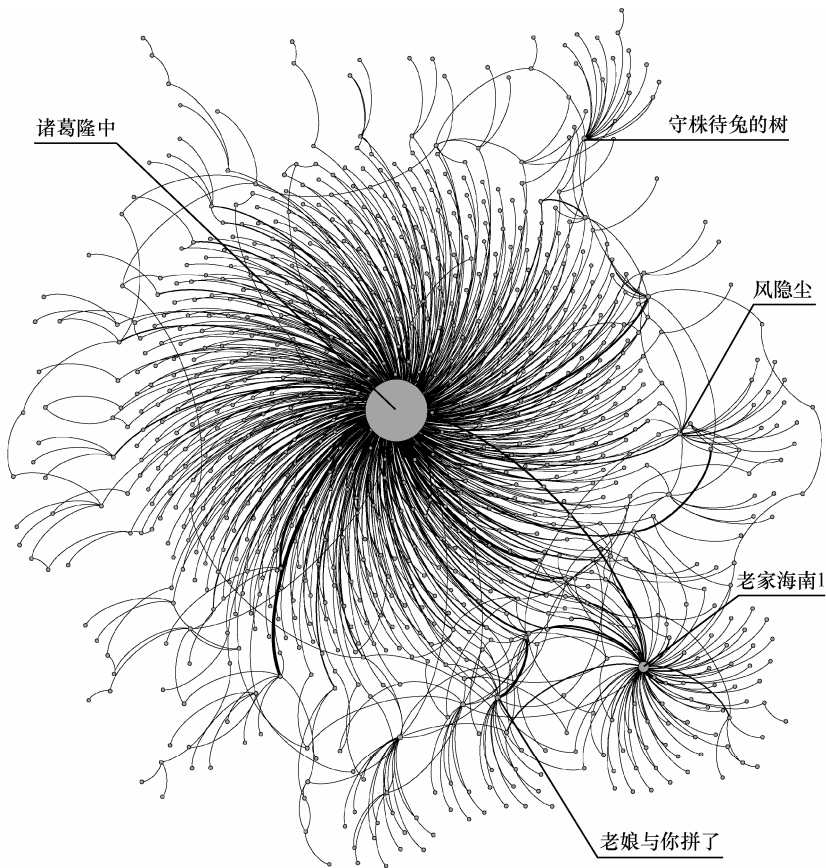


图 8.27 个体网传播图

1) 结构洞

这里列举节点度数排名前 5 位的用户的结构洞指标, 形成相应的数据指标表 (表 8.19)。

表 8.19 结构洞指标

用户	节点度数	有效规模	效率	限制度	等级度
诸葛隆中	662	652.070	0.985	0.227	0.981
老家海南 1	80	79.242	0.991	0.037	0.412
守株待兔的树	18	17.815	0.990	0.074	0.060
老娘与你拼了	18	17.452	0.970	0.163	0.473
风隐尘	15	14.313	0.954	0.142	0.282

从表 8.19 “有效规模” 列看出 “诸葛隆中” 对应的值最大, 达到了 652.070, 这说明 “诸葛隆中” 居于网络的核心。“效率” 列对应值最高的为 “老家海南 1”, 达到了 0.991, 说明该用户在社会网络中和其他节点具有较多的联系, 同时该用户的限制度最小, 说明该用户受到最小的限制。相反, “诸葛隆中” 和 “老娘与你拼了” 的限制度值较大, 说明他们在社会网络中受到较大的限制, 同时他们的有效规模值较小, 说明他们可能仅和某一节点联系。

2) 诚实中间人

从表 8.20 中可以看出, “诸葛隆中” 代表的节点作为中介的次数为 15 次, 共联络了 105 对关系, 其中 105 对关系属于 HBIO 类, 0 对关系属于 HBIO1 类, 0 对关系属于 HBIO2 类。这说明该结点属于纯中介, 参与回复主帖的人物之间的关系比较弱。同时也表明回复是针对主帖回复的, 而几乎没有部分回复针对的是评论用户评论的。

表 8.20 诚实中间人

用户	Size	Pairs	HBIO	HBIO1	HBIO2
诸葛隆中	15	105	105	0	0
老家海南 1	3	3	2	1	0

HBIO 代表纯中介, HBIO1 代表弱中介, HBIO2 代表非中介。

从图 8.27 中可以看出, “诸葛隆中” 发表帖子后, 大量用户参与的讨论, 在回复的用户中, “老家海南 1” 用户发表的评论引起了网民大量回复, 这说明该用户在网络传播中可能会成为一个意见领袖, 他的意见会影响其他用户的思考和回复。

## 8.5 专题 3：国际事故

世界上每天都会发生一些重大事故，有些是人为的，有些是意外。据不完全统计，世界上每天约有上千起事故在发生，每小时都有近千人因为事故而死亡。除发达国家外，像一些非洲国家，很多人每天因为饥饿或战争而死亡。因此，研究国际事故非常必要。

### 8.5.1 专题发展简介

以前，我们从的报纸、新闻等媒体来接触和了解国际大事。随着互联网的普及，转变网上随时随地获取的信息，各种网站的新闻热点的推送，让我们可以多方式多途径地了解国内外大事。所以，现在对国际事故的认知越来越广泛，随之对这些国际事故也越来越关心，并在网络上发起大量谈论。表 8.21 列出 2014 年全年“国际事故”的热帖。

表 8.21 国际事故专题热帖

帖子标题	发帖人	发帖时间	点击量/次	正面回复量/条	中性回复量/条	负面回复量/条	作者情感
马航事件——中国太高明了	二手农民	2014/03/21	650057	922	1627	154	67.215
又是珀斯！2005 年珀斯波音 777 意外事件与马航 MH370 的惊人关联	mrDIC	2014/03/23	802890	545	1012	95	27.731
有几个细节让我一直坚信马航的人还活着——大家看看有没有道理	含章可贞 S	2014/04/21	189787	300	562	48	21.683
马航事件为什么只有我们中国的家属在胡闹？	de52f5d48f8843d-7acf9cc11a16f5556	2014/03/12	513536	2455	4197	894	5.221
马航出事前的“空姐敬告坐飞机的朋友”神帖怎么就沉了	灵界索魂者	2014/03/14	80231	31	1120	21	0.233

如表 8.21 所示，2014 年 3 月 21 日，“二手农民”在天涯论坛上发表了一篇

关于名为“马航事件——中国太高明了”的帖子，引起了广大网民的关注；随后在2015年3月12日，“de52f5d48f8843d7acf9cc11a16f5556”发表的“马航事件为什么只有我们中国的家属在胡闹？”成为了新的热帖。

在帖子的专题传播过程中，大量的网民针对帖子所表达的观点进行了回复，其中也产生了一些颇具影响力的新观点，并引发了网民的关注和讨论。

### 8.5.2 观点分类与观点列表

通过上述的网络数据信息，针对网民在回复过程中的观点挖掘与倾向进行语义的处理，形成了本话题内的核心观点。

#### 1) 观点分类

根据观点涉及的特征词汇和领域以及网民发表的各个观点内容本身之间存在的相关性与差异性，可以将所有评论分为以下三个主题。

##### 主题一：马航两路 MH370 与 MH17 民航飞机失联

此分类观点包含三个关键特征词：马航、MH370、MH17。对应的观点包括：

观点1. 马航民航飞机无故失踪，肯定是大国背后的阴谋。

观点2. 马来西亚给出的解释不能让人信服。

观点3. 马航飞机失联的猜测众说纷纭。

##### 主题二：博拉病毒多国肆虐，美国提供新药运往北非，不收取任何费用

此分类观点包含三个关键特征词：埃博拉、病毒、灾难、伤亡。对应的观点包括：

观点4. 病毒是美国制造出来的，美国趁机再以拯救者身份出现解决问题，无非是自导自演的一场戏。

观点5. 中国应该派医疗队去非洲。

##### 主题三：亚航 QZ8501 失联

此分类观点包含三个关键特征词：施压、政府、波动。对应的观点包括：

观点6. 亚航的失联又是一次美帝的阴谋，这一切都是美帝操纵的。

观点7. 接连民航飞机失联事件让网民认为坐飞机也是非常危险的。

观点8. 亚航的失联是被劫持了。

各主题观点所占比例如图8.28所示。

#### 2) 观点分类详细分析

为了更好地从宏观向微观的深入分析，本节对每个分类观点内的网络用户



分布以及网络民意进行了进一步的分析,通过网络用户真实的反馈信息来反映他们对这一些问题的思考与建议。

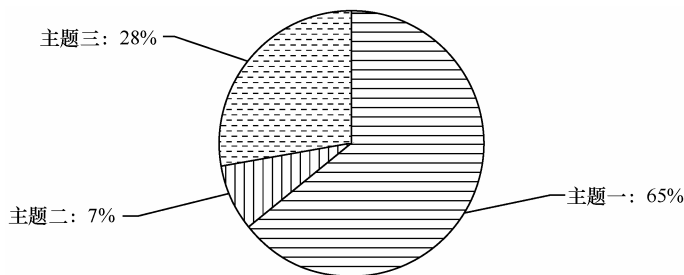


图 8.28 “国际事故”专题观点分类比例

### 主题一：马航两路 MH370 与 MH17 民航飞机失联

此分类观点包含三个关键特征词：升值、衰退、经济。对应的观点包括：

观点 1. 46%，马航民航飞机无故失踪，肯定是大国背后的阴谋。

马航的无故失踪使得人们纷纷猜测，有一部分人认为这是美国的阴谋，为了达到一定的目的发生的事件。

例如，用户“Andrew-M2013”评论说：“天安舰，马航失联，这次的马航击落，对谁有利就是谁做的，这符合制造争端、寻找机会返回亚太战略。”用户“涯天用户”说道：“韩国的民航飞机也是美国为了苏联情报搞的鬼，故意安排飞机偏离航道进入苏联军事重地，才被打下来的，当年美国还亲口承认过。”

观点 2. 29%，马来西亚给出的解释不能让人信服。

对于马航飞机失联之后 24 小时后，马来西亚官方发表了声明，给出了解释，但是说辞太过牵强，明显有违背事实，推卸责任的嫌疑。

例如，用户“笑傲天涯”说道：“马来西亚其实已经发现了客机，也知道整个事情的阴谋，但是碍于大国的威胁，不得不做出不确实的回应。”

观点 3. 25%，马航飞机失联的猜测众说纷纭。

马航失联后，成为了一谜，引起了广泛的猜疑。

猜疑有很多种，用户“法惠生”说：“没有证据证明是美国的阴谋，但是美国长期遏制中国，在非传统安全领域里面的不对等合作，以及对 J 独 Z 独 T 独和脚盆的支持，都造成非传统安全危机。”用户“public20111”猜测：“马航 MH17 被击落的时间是北京时间 7 月 17 日 22 时左右（格林尼治时间 7 月 17 日 14:15），可以看出来，这次凶手依然用航班号作为动手时间，可见两起事件是同一集团所为。而且这一集团的指挥部在欧美，使用的时间是欧美的时间。”

用户“亚天用户”说道：“最大可能是被原苏联导弹 S300 防空导弹击落的，这种导弹在那附近只有乌军和俄军装备有，民兵不会有这么大系统的防空武器。这个倒霉的马航客机应该是被美国 CIA GPS 干扰错误信号导致偏离航线，误进入战区上空，被乌军的 S300 系统击落的，本来乌军要打的并不是该机，而是 17 分钟后通过该附近空域的普京专机，普京躲过一劫，美帝好狠！”用户“对啄 777”评论道：“很明显是乌克兰搞的鬼，上午装模作样设立禁飞区，下午就出事，他们很清楚乌克兰东部势力武装有防空导弹，前几天被打下几架飞机，又故意把民航机往最危险的地方引。”

综述所述，具有相应子观点的评论分布比例如图 8.29 所示。

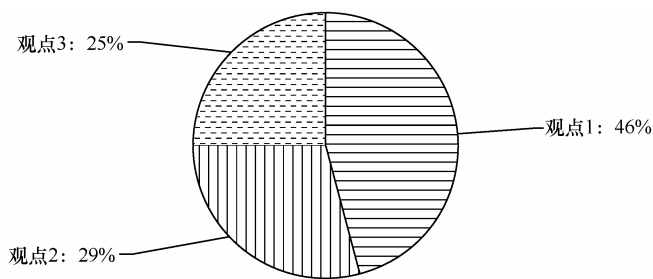


图 8.29 主题一细分子观点比例

## 主题二：博拉病毒多国肆虐，美国提供新药运往北非，不收取任何费用

此分类观点包含三个关键特征词：升值、利益、消费。对应的观点包括：

观点 1. 74%，病毒是美国制造出来的，美国趁机再以拯救者身份出现解决问题，无非是自导自演的一场戏。

对于美国这一慷慨的行为，与之前无利益不美国的做法形成了巨大的反差。所以说，当得知这一消息后，很多网友都对美国的行为表示了怀疑。

例如，用户“谁的人生不是唯一”说：“据说美国治疗埃博拉病毒的解药来自于一个仅有 6 个人的新成立的公司。这是先有病后有药，还是先有药后有病啊？傻傻的，分不清。反正感觉美帝很厉害：6 个人成立一个新公司迅速就拿出了解药。”用户“mechalcao”评论说：“可以免费的临床试药，还不用担心什么后果，这种天大的好事谁不做就是傻子！要在国内试，还得花钱请人，万一出点纰漏，还会被舆论鞭打，被政府找麻烦！”此外，用户“我思故我在 123ABC”评论说：“这病毒本来就是美国制造的。美国在制造病毒之前，就已经研制好了疫苗。这背后的意图，一看就清楚。”

观点 2. 26%，中国应该派医疗队去非洲。

美国宣布这消息之后，不论他们真实目的是怎么样的，但是却打着正义救援的名义；作为一向以和平、友好为自己的外交方针的国家，我们也应该在西非国家树立好的旗帜。

例如，用户“小盐平八郎”说：“我国政府怎么不派医疗队去非洲？”用户“类比星”说：“对于医疗条件那么差的非洲，我们应该提供我们应有的帮助。支援非洲，彰显我大国雄风。”

综述所述，具有相应子观点的评论分布比例如图 8.30 所示。

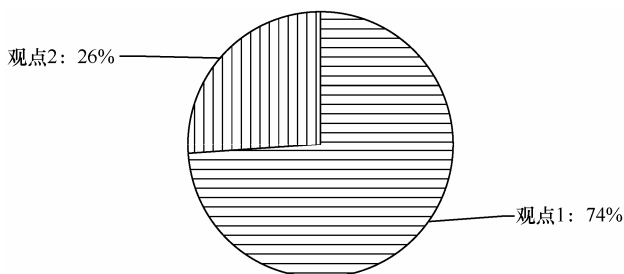


图 8.30 主题二细分子观点比例

### 主题三：亚航 QZ8501 失联

此分类观点包含三个关键特征词：施压、政府、波动。对应的观点包括：

观点 1. 52%，亚航的失联又是一次美帝的阴谋，这一切都是美帝操纵的。这个观点认为和马航一样，亚航的失联一定又是大国为达到军事目的的牺牲品。

例如，用户“万仞千山一肩负”说：“丐帮苏灿的那个帖子中有人隐约有爆料，说是亚航事故原因其实已经调查清楚了，原因很可笑，但两国为了政治利益都隐瞒了真相。参与搜救的各国都知道内幕，结果都隐瞒了真相。真是屁民不知情权啊，不管在哪国都一样。”用户“inferno79”说：“金融是现在西方国家剥削其他国家的核心手段，其他方面跟西方叫板都可以商量，金融上谁叫板谁死。中国这样的大块头出来牵头叫板死不了，可想跟着中国玩的小国就包死得挺挺的。喷完老美喷国内，小贷公司的放开可以说是近几年最失败的政策，不但不能缓解企业融资难的问题，反而更加推高了融资成本。”

观点 2. 33%，接连民航飞机失联事件让网民认为坐飞机也是非常危险的。接连几次飞机失事，并且被证明乘客无一幸免，这件事情对论坛的用户产

生了巨大的影响，很多用户表示不敢坐飞机了。

例如，用户“飞逝的九九”说：“这世道太乱了，连飞机都不安全了。”用户“大块得”说：“再也不信爱了，世界太乱，我还是在家吧……”

观点3. 15%，亚航的失联是被劫持了。

在众多猜测中，有的用户怀疑亚航被劫持了，是恐怖分子的作为。

例如，用户“河北老李 ABC”说：“亚航飞机是被人劫机了，不然怎么得不到求救信号呢。”用户“911 犹太集体请假”说：“亚航刚出发 8 分钟就失联了，毫无踪迹可言，所以必定是发生了突发性事故。”

综述所述，具有相应子观点的评论分布比例如图 8.31 所示。

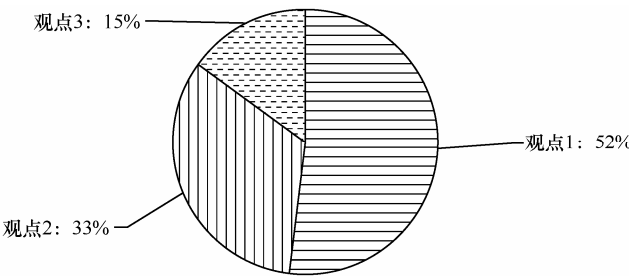


图 8.31 主题三细分子观点比例

### 3) 观点总结

从网友对于“国际事故”问题的讨论中可以看出，对于众多国际事故，网民对马航事件的关注度比较大。国际事故是灾难的起源，世界上每天因灾难事故死亡的人不计其数，如地震，车祸，自然灾难等。这些事件也都发生在我们身边，所以关注国际事故，提早预防，增加事故灾难知识，才能更加美好地生活。

### 8.5.3 专题传播趋势

本节通过对专题影响力和专题情感两个方面的研究来表征专题的传播趋势。专题影响力分析的基础是专题中帖子的有效回复数量，专题情感分析的基础是帖子回复内容的情感趋向，涉及情感计算。

#### 1) 影响力

本节分别对新增回复数和总回复数按照时间进行统计，然后对统计结果进行可视化得到波线图（图 8.32，图 8.33），以便发现专题传播的发展峰值。

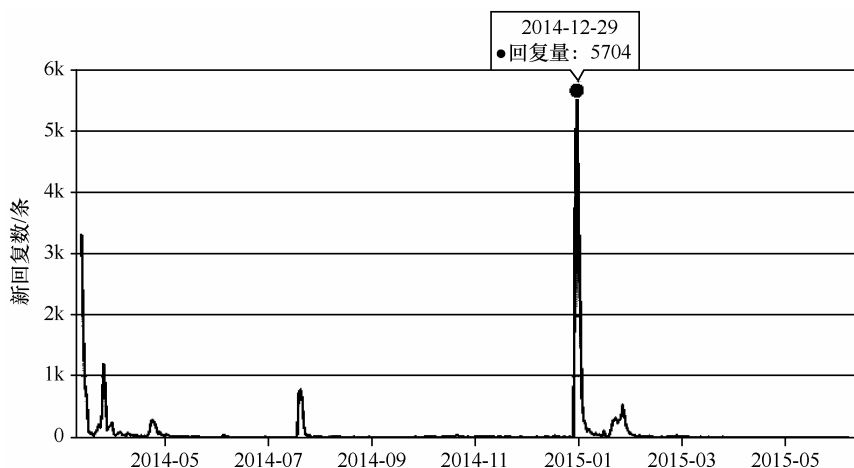


图 8.32 专题日流量时间趋势图

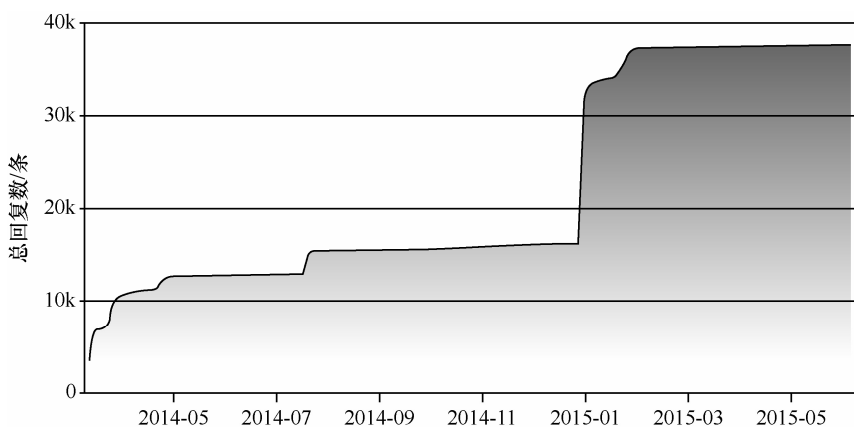


图 8.33 专题总流量时间趋势图

从图 8.32 可以看出每天对应的新增回复数的变化情况。最高点对应的日期为 2014 年 12 月 29 日。2014 年 12 月 29 日之前，每天对应的新增回复量非常少，说明这一段时间该事件提出来之后并没有引发大的影响；而在 26 日之后，每日新增量开始急剧上升，并且仅仅用了 3 天时间，在该月的 29 日达到了发展峰值，说明该专题在这三天引发了激烈的讨论。在 29 日之后讨论急剧下降，以后基本话题讨论就几乎没有。由该事件的时间特点可以看出这是一个马航事件，5 月 12 日左右马航失联，在 12 月 29 日左右的亚航失联又将马航事件引出，两者对比，所以对应的话题量达到了一个峰值。

## 2) 传播情感分析

本节利用 HowNet 建立一个有效的中文情感词库，对专题内的回复内容进行了情感分析。情感分析分别对情感趋势和情感分类的统计结果进行可视化得到情感总体趋势图（图 8.34）和专题情感分类图（图 8.35）。

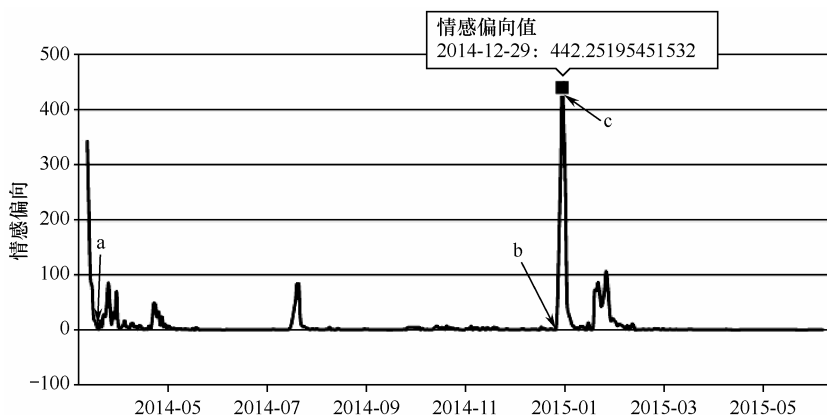


图 8.34 专题情感趋势图

点 a：2014 年 5 月 17 日；点 b：2014 年 12 月 26 日；点 c：2014 年 12 月 29 日

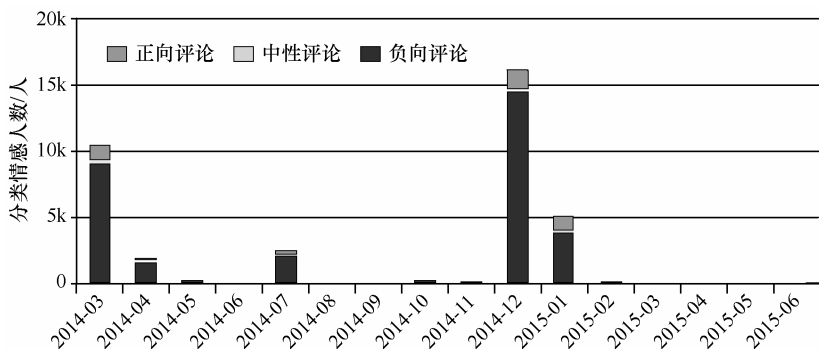


图 8.35 专题情感分类图

从图 8.34 可以很容易看出发展峰值和发展拐点。点 b 对应的情感偏向的情感值比最高，并且达到峰值；点 a 对应的情感达到较低的点，接着达到峰值，说明这写点属于情感拐点，必定有一些事情影响了网民的情感。

从图 8.35 可以看出在网民的评论中，负面评论占的比重较大，说明现在网络环境下，喷子较多，网民发言还是比较随意，大部分网民选择在网络上发泄自己的情绪。

### 8.5.4 专题用户情感分析

探索网民情感可以发现情感积极的用户，可以找出产生消极影响的网民，对网络舆情监控具有重要意义。

本节对各专题下的人物情感进行了统计分析，由于活跃度对情感的影响较大，所以本节分别以情感值和活跃度为两个坐标轴构建专题任务散点图（图 8.36）。

从图 8.36 中可以看出，一般发帖人的情感值都比较高一些，向那些圆点的情感值比较高况且活跃度也很大，所以这些用户很可能是意见领袖。但是也存在一些点不是发帖者但是其活跃度较高，说明有两种可能，一种是该用户作为评论人，评论的内容又引发了一个新的讨论热潮，另一种是属于那些发帖较少，经常参与那些评论或者原帖回复的那一类人，这些人有一定的回复参与量，但是通常回复的情感极性较低，感情色彩不明显，但是这些人所属领域性比较明显。

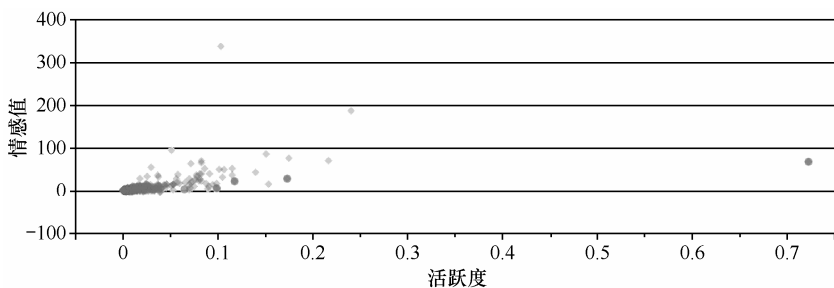


图 8.36 基于活跃度和情感值的用户散点图

圆点代表发帖人，方点代表回帖人

为了进一步进行分析，本节将那些情感值负性交大并且活跃度较小的人进行了详细点的描绘，取情感值小于 20，活跃度大于 0.02 的所有用户，得图 8.37。这类用户代表的是那些在论坛活动积极，但是负面情绪较大的那些人。这些人的发言或者参与会影响到人们对于事件情感的判断。由于他们的活跃度相对来说较大，所以他们影响其他用户的可能性比较大，会给论坛用户灌输一些负面的情绪。因此，很容易从这类人群中发现一些用户，并且对其进行权限的限制、屏蔽、删除。

图 8.37 箭头所指向的点分为两类：第一类点是标出具体指标的点，该用户活跃度比较大，但是他的情感属于最低的（-5），这类用户极有可能是负面情绪的发起者，论坛管理员应该采取相应的操作；第二类点是图中标出的另

一个点，该用户活跃度非常高，比其他用户高出很多，但是其情感值却只在 1 附近，这类用户极易被利用，发展成为消极情绪的发起者，应予以关注。

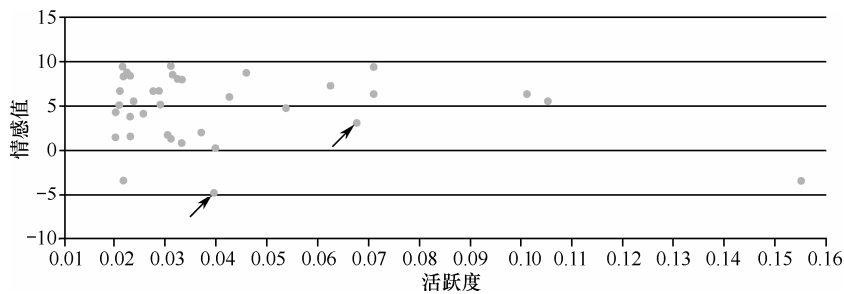


图 8.37 低情感和高活跃度散点图

8.5.5 专题网络的构建与分析

构建专题网络的方法见节 4.3.5。构建国际事故的专题网络，进行可视化结果见图 8.38。

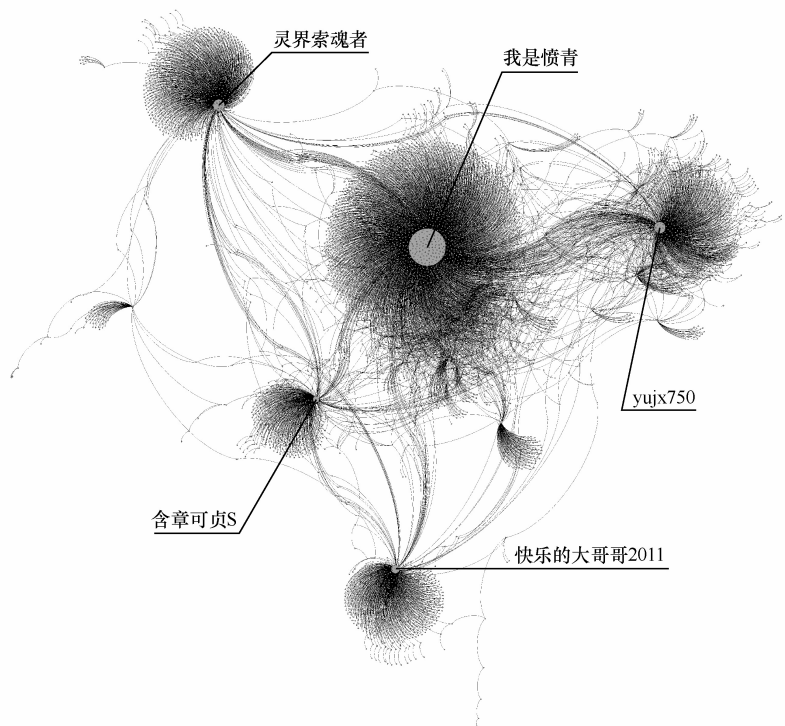


图 8.38 专题传播网络



从图 8.38 中可以明显看到,“我是愤青”和“灵界索魂者”的两个用户发表的帖子引起论坛用户广泛的讨论。在整个整体网中,节点之间的联系都比较紧密,表明用户在该主题下讨论的问题都比较集中。除了前面两个最大的社区之外,“yujx750”“快乐的大哥哥 2011”“含章可贞 S”“步雨中”相对节点度数也较高,他们的观点和意见影响了相当一部分的用户。

1) 网络指标概述

本节分别从社会网络的整体性指标(表 8.22)、中心性与影响力指标、关联性与“小世界”指标进行了分析,以便发现网络中的特殊节点,以便分析个体网络演化与传播。

表 8.22 整体基本指标

平均度	图密度	网络直径	模块化指标	平均路径长度	平均聚类系数
2.477	3.305	12	0.618	3.428	0.482

平均度:较小,平均加权重为 2.477;图密度:该传播网络图比较稀疏;网络直径:较长,该网络的传播较远;模块化指标:划分出 7 个社区,且划分的质量较高;平均路径长度:较大,网络传播较广;平均聚类系数:聚类效果不错。

2) 中心性与影响力

使用 UCINET 对该网络的中心度指标进行计算,这里取点度中心度排名前 5 的节点,列出这些节点的点度中心度、中间中心度和接近中心度三项指标的绝对值、相对值和排名,得到对应的数据如表 8.23 所示。

表 8.23 中心度排名

用户	点度中心度			中间中心度			接近中心度		
	绝对值	相对值	值排名	绝对值	相对值	值排名	绝对值	相对值	值排名
我是愤青	5197	0.540	1	4704	0.008	69	20371156	0.037	1
yujx750	1540	0.160	2	0	0.000	3546	40023420	0.019	4
灵界索魂者	1080	0.112	3	0	0.000	6418	39776320	0.019	3
快乐的大哥哥 2011	780	0.081	4	0	0.000	933	42409832	0.018	5
含章可贞 S	736	0.077	5	136232	0.245	1	48028936	0.016	77

从表 8.23 中可以看出“我是愤青”“yujx750”等 5 个节点的点度中心度排名前 5 位。从中间中心性来看,“含章可贞 S”的中间中心度远远大于其他节点,排在第一位,说明许多节点都需要经过该节点才能建立联系并获取信息。从接

近中心性来看,“我是愤青”排在第一位,说明该节点与其他节点进行交互比较容易。

图 8.38 的点度中心势指标为 0.54%,该指标值很小,表示该图在度数上向某点集中的趋势很不明显;该图的中间中心势指标为 0.24%,该指标值也比较小,表示该图向某些中间节点聚集的程度比较小,即在整个网络中大部分的结点并不需要经过某些中间节点,就能得到信息。

3) 关联性与“小世界”

图 8.38 的特征途径长度为 5.018,凝聚度为 0.002。该图路径长度的具体分布如表 8.24 所示。

表 8.24 路径分布指标

路径长度	出现频次	占比	路径长度	出现频次	占比
1	9496	0.025	9	11485	0.030
2	42009	0.110	10	5797	0.015
3	40159	0.105	11	3001	0.008
4	60923	0.160	12	1213	0.003
5	83951	0.220	13	477	0.001
6	63110	0.165	14	137	0.000
7	37772	0.099	15	42	0.000
8	22110	0.058	16	6	0.000

图 8.38 的聚类系数为 3.749,加权的聚类系数为 0.002。节点聚类系数如 8.25 表所示。

表 8.25 节点聚类系数

用户	聚类系数	点对数
我是愤青	0.000	6402831
yujx750	0.000	522753
灵界索魂者	0.000	524800
快乐的大哥哥 2011	0.000	252405
含章可贞 S	0.000	120786

8.5.6 个体网络分析

按照整体网的分析之后,选择一些传播特性相对较强的用户做一个帖子下所有用户的个体传播网络。本节选取“国际事故”发表的一篇帖子,对它进行

指标的分析，并且描述传播路线。图 8.39 为通过用户发表的帖子中用户的评论关系构建的传播网络。

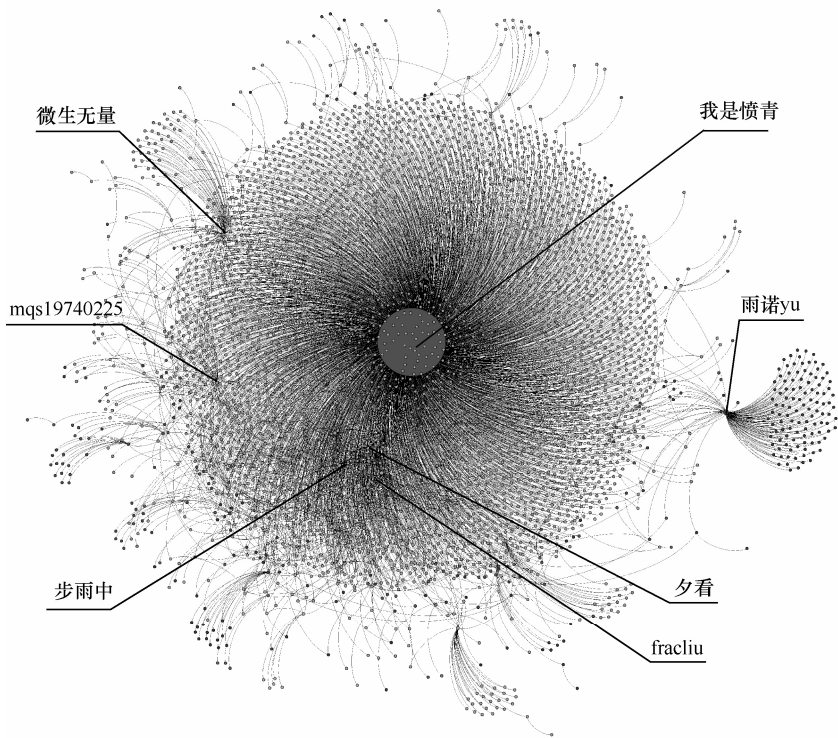


图 8.39 个体网传播图

1) 结构洞

这里列举节点度数排名前 5 位的用户的结构洞指标，形成相应的数据指标表（表 8.26）。

表 8.26 结构洞

用户	节点度数	有效规模	效率	限制度	等级度
我是愤青	3580	3406.323	0.951	0.220	0.985
雨诺 yu	134	133.849	0.999	0.012	0.232
微生无量	84	83.264	0.991	0.067	0.602
步雨中	71	53.516	0.754	0.355	0.700
fracliu	64	51.663	0.807	0.318	0.683

从表 8.26“有效规模”列看出“我是愤青”对应的值最大，达到了 3406.323，

这说明“我是愤青”居于网络的核心。“效率”列对应值最高的为“雨诺 yu”，达到了 0.999，说明该用户在社会网络中和其他节点具有较多的联系，同时该用户的限制度最小，说明该用户受到最小的限制。相反，“步雨中”和“fracliu”的限制度值较大，说明他们在社会网络中受到较大的限制，同时他们的效率值较小，说明他们可能仅和某一节点联系或少量节点有联系。我们猜测这类点代表的用户很可能在社区间起桥接作用。

2) 诚实中间人

从表 8.27 中可以看出，“夕看”代表的节点作为中介的次数为 13 次，共联络了 78 对关系，其中 72 对关系属于 HBIO 类，2 对关系属于 HBIO1 类，4 对关系属于 HBIO2 类。还可以看出纯中介所占比例最大同时弱中介和非中介也有一定的比例。这表明大部分回复是针对主帖回复的，而其他部分回复针对的是评论用户评论的。

表 8.27 诚实中间人

用户	Size	Pairs	HBIO	HBIO1	HBIO2
夕看	13	78	72	2	4
mqs19740225	12	66	65	1	0
Fracliu	12	66	59	6	1
步雨中	9	36	32	1	3

HBIO 代表纯中介，HBIO1 代表弱中介，HBIO2 代表非中介。

从图 8.39 中可以看出，“我是愤青”发表的帖子后，大量用户参与的讨论。在回复的用户中，“雨诺 yu”“微生无量”“步雨中”等用户发表的评论引起了网民大量回复，其中“雨诺 yu”和其他用户的互动次数明显更多，说明该用户在网络传播中可能会成为一个意见领袖，他的意见会影响其他用户的思考和回复。

8.6 本章小结

本章以天涯论坛的国际与军事领域舆情研究领域，分别从领土主权、合作共荣和国际事故三个专题进行了研究和分析。通过分析，得到了一些相关的网络舆情，包括网民对国际与军事主要问题的一些观点以及情感态度。

同时，通过构建国际与军事领域整体网、高频词时间演化网、专题传播网络等社会网络，应用社会网络分析法，从而直观地发现和展示了一些

与内容分析相符的结论。从本章节我们发现，网民对国际上的大事非常的关注，尤其是对那些和中国相关的事件，例如马航事件，钓鱼岛问题，南海问题等，我们网民都有比较高的参与积极性，并且对那些事件发表自己的不满和愤怒，说明中国公民有很强的民众责任感，另外，网民对那些国际上的灾难事故也十分关注，从情感值的正负倾向来看，网民十分关心事故灾难所带来的伤亡与伤害，说明中华民族又是一个心系天下的民族，极富有同情心和援助意愿。从这些方面，探索这些舆情热点可以间接体现出中华儿女的热血豪情。

## 第9章 社会舆情分析工作总结与展望

自从党的十八大召开以来，我们国家内政与外交发生了许多新的变化，在人民生活稳步提高的同时，各个领域的建设也在快速发展，综合实力不断增强。但是，随着社会与经济的快速发展，也存在一些因管理机制和体制而产生的一些社会矛盾，这些矛盾通过网络中存在的舆情所反映出来，也迫使我们进一步加快改革的步伐，通过不断地深化改革，来加强我国的全面建设。

因此，我们希望探讨国家在各领域建设中所发生的热点事件，并从其中反映出网民们所关注的问题与建议，以及网民参与讨论过程中对这些问题的看法、情绪以及期望，利用网络所获取的真实数据进行深入的分析与研究，尝试对这些进行定性以及量化的分析，为更多的民众提供一个数据参考。并且，我们希望利用具体的实际问题的研究，促进我们在舆情分析领域的技术进步，特别是在文本挖掘、网络情感分析以及基于结构的社会网络分析法等领域取得新的进展，这也是研究与制定本报告基本出发点之一。另一方面，也希望相关结果可供更多领域人士一起来研究和探讨，特别是希望从另外一个角度——网民的角度来看世界的变化以及对热点事件的反映；如果将网民舆论中存在的理性分析与智慧以及民意信息收集起来，并在未来能够应用到单位或者政府机构的政策制定过程之中，这将会成为本报告的另一重要的价值。

我们对 2014 年 1 月到 2015 年 3 月期间，天涯论坛上涵盖经济、政治、社会、文化、生态及军事等领域近 75 万条帖子及其构成的热点专题进行文本数据的过滤、特征抽取、建立指标与模型、进行算法的优化与处理等一系列过程，得出 2014 年的热点事件列表，并利用舆情内容挖掘和社会网络结构分析方法相结合，对这一些事件进行分析和总结。

### 9.1 社会舆情分析总结

经济领域中，经过数据过滤与清洗后，针对 2014 年内发布到天涯论坛中有关经济领域的 11000 多条帖子进行高频词分析，发现经济领域内的帖子大体上可以分为“房市与房价”“股市与股票”“货币与汇率”三大主题，因此我们分别构建了三大主题的主题词库，应用该主题词库进行分类，并对每个领域内

的主题进行舆情分析。通过对这些专题中网友发布的帖子和回复的分析可以得出,控制房价上涨成为2014年度此领域中最受网友关注的话题,控制房价上涨、预防房价泡沫成为政府需要投入大量精力的重要工作领域。在股市与股价方面人们关注的不是具体的股票的增长或者跌落,而是大部分关注点移向了炒股的技巧和心态上面,特别是在牛市行情中,好的技巧和心态会帮助股民更好地适应股票市场,这也反映中国股民在投资心态上的逐步成熟。另外,广大网民针对人民币升值和汇率变化带来的问题和影响进行了深入持久的讨论,汇率的波动事关每个人的利益,在关注汇率波动、分析波动趋势的同时,建议采用灵活调整汇率的制度来保证大众利益不会受到侵害。同时,也反映了大量网民对国家金融政策具有较强的信心和较高的信任度。

政治领域中,经过数据清洗与过滤后,针对2014年内发布到天涯论坛中有关政治领域的近10000余条帖子进行高频词分析,发现政治领域内的帖子大体上可以分为“国企改革”“反腐风暴”“依法治国”三大主题,在构建三大主题的主题词库后,利用该主题词库进行分类,并对相应领域内的热点主题进行舆情分析。通过对这些专题下网民帖子、回复的分析可以得出,改革是一项持续性工作,在进行政治体制和制度改革过程中,需要维护法律权威,确保依法独立行使审判权、检察权,在完善司法保障制度的基础上,系统化与制度化的去解决违法犯罪、贪污腐败等社会热点问题。同时,通过推进政府治理体系和治理能力的现代化,切实转变政府职能,把政府职能转向更好维护社会的公平正义,创造更好的发展环境,提供更加优质的公共服务等方面。

民生领域中,通过数据清洗与过滤后,针对2014年内发布到天涯论坛中有关经济领域的13000多条帖子进行高频词分析,发现民生领域内的帖子大体上可以分为“教育与就业”“收入分配”“社会与医疗保障”三大主题,并分别构建三大主题的主题词库,应用该主题词库进行分类,分别对每个领域内的主题进行舆情分析。通过对这些专题下网友帖子、回复的分析可以得出,“教育”是强国富民的基础,“就业”是人民生存和生活的根本,教育资源分配不均、中国教育的现状、大学生就业和创业问题是教育与就业领域的热点话题。“分配”是人民休养生息的源泉,就是“改革发展成果让人民共享”,公务员薪资问题、教师待遇问题以及中国高物价与低工资的矛盾涉及社会的分配体系的公平性。“社保”是人民生存和发展的依托,工人退休问题以及当前的社会和医保现状也引起了网友的热切关注。构建和谐社会,就是要把民生问题作为重中之重,让广大人民群众有活干,有学上,有饭吃,有衣穿,有屋住,病有医,

老有养，生活幸福，都过上好日子。可见，民生问题不但事关广大人民群众的根本利益，而且影响整个国家改革发展的大局，将发展民生的目的真正体现到满足人民需要、实现人民利益、提高人民的生活质量与幸福指数上，这也是党中央工作的重中之重。

生态领域中，经过数据清洗与过滤后，针对 2014 年内发布到天涯论坛中有关生态领域的 40000 余条帖子进行高频词分析，发现生态领域内的帖子大体上可以分为“生态环境”“生态资源”“技术与措施”三大主题，并构建三大主题的主题词库后，得用该主题词库进行分类，分别对相应领域内的主题进行舆情分析。通过对该领域中网民发布的帖子和回复的分析可以得出，建设美丽中国深化生态文明体制改革过程，要从污染源头开始考虑，系统化地分析可能产生的原因与影响，从而科学地进行决策，如备受网民关注的雾霾问题。此外，在采用新型生物技术提供资源的同时，也要有一定的前瞻性，对于未来生物进化产生影响较大的问题，如“转基因育种”等，需要进行制度决策与技术预研等协调过程。通过加快建立生态文明制度，健全国土空间开发、资源节约利用、生态环境保护的体制机制，推动形成人与自然和谐发展的现代化建设新格局。

国际与军事领域中，通过数据清洗与过滤后，针对 2014 年内发布到天涯论坛中有关国际与军事领域的筛选后的 6000 多条帖子进行高频词分析，发现国际与军事领域内的帖子大体上可以分为“领土主权”“合作共赢”“国际事故”三大主题，分别构建三大主题的主题词库，应用该主题词库进行分类，分别对每个领域内的主题进行舆情分析。通过对这些帖子进行聚类分析，得到最热门的专题是“钓鱼岛的所属权问题”，通过对该专题中网民发布的帖子和回复的分析可以得出，绝大多数网民认为钓鱼岛是中国的，日本依仗美国势力抢占钓鱼岛的所属权。同时，网民认为主要的原因在于日本不负责任的言论与做法极大地影响了两国之间的关系，导致两国民众的对峙局面。而在 G7 峰会上日本提出并批评中国在南海问题上的做法，这一事件在近两年的论坛中持续引起了网民广泛而激烈的讨论，网民们认为这是日本借助 G7 强国的势力来向中国施压，在国际上制造中国威胁的舆论。而网民对于“马航失联事件”的猜测层出不穷，大量的舆论认为可能是一些帝国主义势力为了实现自己目的而制造的人为事件。综上所述，伴随着中国的强大以及国际地位的提升，一些国家可能还会针对中国而制造出一些舆论或事件来达到遏制中国发展的目的，所以关于中国周边的大量事件成为今年舆论的热点。因此，我们也应该在坚持主权和领土完整的基础上，探索维护我国国家以及公民的合法权益的有效方式。



## 9.2 下一步工作展望

基于大数据的分析已成为网络舆情分析以及热点分析的核心技术。大数据技术是指运用搜索引擎、社交媒体、各类网络数据库，实时聚集数以 TB 级存储量的文本和图像等数据资源，利用海量数据存储与索引、文本分析、特征抽取、数据聚类与分类、算法优化以及数据可视化等技术为一体的综合性的应用技术。通过这样的技术，可以精确描绘现状并预测未来。大数据正在变革我们对政治、新闻、商业、健康、教育的传统认识。

西安交通大学软件学院社会智能与复杂数据处理实验室，致力于利用不同数据来源的各类数据，包括结构化数据、非结构化数据、图像和语音等媒体数据，在特定的应用场景中进行数据分析、挖掘与深度学习。一方面，我们力图将最有效的算法与技术应用于解决实际中出现的问题；另一方面，我们也希望利用创新性技术手段解决更多的未知问题。目前，实验室正在建立统一的中文情感字库、领域词库，深入研究本体以及本体映射机制，数据流的实时聚类与演化分析，以及数据可视化等技术的深入研究与应用实践。

本报告就是在这样一个背景基础下开始提出并完成的。由于在有限的时间约束下以及关键技术攻关等问题的影响下，本报告暂时只针对了一种单一的数据源和有限的数据展开了相关研究和分析，尽管其中反映了一些有价值的结果，但是与我们的未来目标相差很远。我们也希望在不远的将来，可以将技术与应用紧密结合，对多数据源以及更多的数据类型在复杂条件下进行实时的数据分析。另外，我们也希望在未来的 10 年中，在同行各位专家的指导下，通过深入交流与合作，可以不断地升级我们的分析数据的理论、方法与技术，从而形成一个更合理、更准确、更自动化的分析工具。通过这些工具可为社会提供更多有意义的数据分析与知识服务，这也是我们下一步的工作重点。

## 主要参考文献

- 陈京民, 韩永转. 基于虚拟社会网络挖掘的网络舆情分析[J]. 中国制造业信息化, 2010, 39(5): 65-67, 71.
- 陈然. 网络舆论领袖测量方法初探 [EB/OL]. [2011-01-15]. <http://www.tsjc.tsinghua.edu.cn/newmedia/wordpdf/10.Pdf>.
- 程倩. 网络舆情模型构建与网民分析[D]. 郑州: 郑州大学硕士论文, 2011.
- 戴媛. 我国网络舆情安全评估指标体系研究[D]. 北京: 北京化工大学硕士论文, 2008.
- 第 33 次中国互联网络发展状况统计报告. 中国互联网络信息中心网 [DB/OL]. [2014-03-05]. [http://www.cnnic.net.cn/hlwfzyj/hlwxyzbg/hlwtjbg/201403/t20140305\\_46240.htm](http://www.cnnic.net.cn/hlwfzyj/hlwxyzbg/hlwtjbg/201403/t20140305_46240.htm).
- 高洪杰. 互联网舆情监测分析系统实现[D]. 上海: 复旦大学硕士论文, 2009.
- 郭瑞涛. 中国网民主体特征分析[D]. 长沙: 湖南师范大学硕士论文, 2012.
- 黄雯, 王旗. 大气颗粒物化学成分与健康效应的关系及其机制的研究进展[J]. 卫生研究, 2012, 41(2): 323-327.
- 黄晓斌, 赵超. 文本挖掘在网络舆情信息分析中的应用[J]. 情报科学, 2009, 27(1): 94-99.
- 蒋立峰. 未来十年的中日关系与中国对日政策——21 世纪中日关系研究报告[J]. 日本学刊, 2009, 5: 3-17.
- 柯惠新, 刘绩宏. 重大事件舆情监测指标体系与预警分析模型的再探讨[J]. 现代传播, 2011, 185(12): 39-44, 56.
- 李雯静, 许鑫, 陈正权. 网络舆情指标体系设计与分析. 情报科学, 2009, 27(7): 986-991.
- 李稳国. BBS 信息传播网络的分析与研究[D]. 南宁: 广西师范大学硕士论文, 2008.
- 林敏. 网络舆情: 影响因素及其作用机制研究[D]. 杭州: 浙江大学博士论文, 2013.
- 刘军. 社会网络分析导论[M]. 北京: 社会科学文献出版社, 2004, 4-5.
- 刘毅. 网络舆情与政府治理范式的转变 [J]. 前沿, 2006, (10): 140-143.
- 罗家德. 社会网络分析讲义[M]. 北京: 社会科学文献出版社, 2005, 171-175.
- 齐建国. 解决雾霾问题应从社会经济角度切入[N]. 中国经济导报, 2013-02-02.
- 谈国新, 方一. 突发公共事件网络舆情监测指标体系研究[J]. 华中师范大学学报(人文社会科学版), 2010, 49(3): 66-70.

- 王宏勇. 网络舆情热点发现与分析研究[D]. 成都: 西南交通大学硕士论文, 2011.
- 王来华. 舆情研究概论——理论、方法和现实热点[M]. 天津: 天津社会科学院出版社, 2003.
- 王青, 成颖, 巢乃鹏. 网络舆情监测及预警指标体系研究综述[J]. 情报科学, 2011, 29(7): 1104-1108.
- 约翰·斯科特. 社会网络分析法 [M]. 刘军, 译. 重庆: 重庆大学出版社, 2007.
- 张一文, 齐佳音, 方滨兴, 李欲晓. 非常规突发事件网络舆情热度评价指标体系构建[J]. 情报杂志, 2010, 29(11): 71-75, 117.
- 张瑜. 基于评论性网站用户发言数据挖掘的研究[D]. 北京: 北京邮电大学硕士论文, 2013.
- 张玉峰, 王志芳. 基于内容相似性的论坛用户社会网络挖掘[J]. 情报杂志, 2010, 29(8): 125-130.

## 后 记

子在川上曾曰：“逝者如斯夫”。时间真快，从开始写这一本书的前言，到写后记时正在经历了一次时空的大挪移，一方面，我们的实验室正在迎来三周岁的生日，三年的努力与辛苦的岁月如歌，也使我们培养出了两批优秀的学生，其中大部分人已分布在中国 IT 行业的一线，而参与本书的主要实验设计与撰写任务的同学们也将面临新的就业机会；另一方面，我个人也在从西安交大软件学院 SICDP 实验室来到大洋彼岸的卡内基梅隆大学 LTI 的 Informedia 实验室进行适应的过程中，深夜无眠，利用“倒时差”的工夫，来完成了这一篇后记。

作为我们实验室定期发布的系列研究报告中的第二本正式的书籍，本书本身就体现出了两个方面的事情，即继承与创新。一方面，我们希望可以在已有工作的基础上，将原有的分析方法、指标以及算法直接可用；另一方面，我们更希望能够站在原来工作的基础上，将社会化网络分析方法、文本挖掘算法以及指标体系融入到社会舆情与信息传播的过程之中，通过对这些信息的处理与分析，使我们可以站更高的角度来审视原来本已习惯的、不被人们特别所关注的的数据信息。因为，人们常说数据是“金”，如何利用有效的手段来挖掘出数据中的价值是所有人的期望和要求。自然对于希望从事对于社会化网络中的复杂数据研究的我们而言，更是如此。

为了更为深入地开展工作，我们这一次报告还是将论坛的数据作为研究与分析的对象，一方面是由于数据来源的开放性与完整性，我们可以更快地获取到网络中全年度的相关信息；另一方面由于网民进入的门槛更低，从而为我们深入研究社会各阶层网民，特别是“草根网民”的民意与民声提供了重要的舆情数据来源与分析研究基础。但是，在移动互联网的时代，社交媒体的应用模式、传播机制以及分析的实时性需求已经发生了重大的变化，这均对我们未来的研究提出新的应用需求和技术挑战。

自从戴汝为院士提出社会智能这一个概念以来，网络技术、用户使用模式、网民数量以及网络信息数据总量均已发生的巨大的变化，通过网络与数据，实现了从“集成的演练厅”的精英化决策模式向“群体智慧”的网络群体决策模式的演化。也许在大数据时代下，社会智能将会成为社会科学与计算机软件工程交叉融合一个重要的发展方向，它会将物理社会的真实信息转移到互联网平台之上，从而率先实现社会群体信息与行为的可计算，在此基础上，人们利用互联网进行交友和组织活动、网络商城中购买商品与评价、在线学习与反馈、

网络金融与在线炒股，甚至可以通过网络来实现企业竞争情报分析以及国土安全与防恐怖分析的国家信息安全战略，在这些具体和明确的应用需求的推动下，利用网络信息来实现对社会活动与社会业务的智能化分析与决策将会已成为信息领域的国家级的重点研究和发展领域。

另外，网络技术的发展也改变了人们对软件与管理的认识，人们越来越多地从服务的角度来再思考软件与平台的价值，以及信息系统与信息内容的价值，因此，我们在完成研究报告的同时，也在不断地调整和优化我们开发的软件平台，也希望平台的开发，将会直接促进对网络舆情需求的内容服务模式的产生。这必然会对我们的工作提出了一个新的挑战，即利用内容挖掘的技术与方法，特别是文本挖掘与自然语言处理技术，来实现对内容知识的抽取、摘取生成、知识聚合以及知识服务全生命周期的服务体验，这也是我们希望未来可以不断完善与改进的新的研究领域和方向。

这一些技术的挑战下，使得我们自己很容易找到自己的未来方向以及目前的不足，无论是指标体系的建立、算法的优化、辅助可视化工具的实现以及信息与知识服务的深度研究，均需要找到一个阶段的里程碑事件，而本书的出版，将代表了这一个阶段中，我们对研究工作的初步结果；同时，也将会成为下一个阶段的研究基础。在此基础上，可以不断地优化并形成特定主题的文本语料库，利用专业的软件工具以及数据挖掘的核心算法，建立起一套完整的数据分析指标体系，并实现数据分析的自动化，并为研究在特定语境条件下的中文分析与文本挖掘的算法奠定基础。

在报告准备提交的过程中，曾与科学出版社签订了合作协议，但人事的变更与规则的调整，导致我们不得不选择更具有长远战略眼光的一家出版社，希望在日后的合作中，共同建立一个品牌，并形成相对长远的一个学术影响，我们也会致力于此来长期努力。感谢电子工业出版社的甄文全编辑，作为在本书出版的原始合作人，一路上有许多问题可以通过简单、明了和高效的方式来得到沟通解决，甄总的大度与眼光也是我们建立长期合作的基础。同时，也感谢秦绪军社长热情协助与支持，希望通过长期的稳定合作来共同建立起本系列书籍的品牌与社会影响度。

同上本书一样，为了避免一些不必要的断章取义的行为，我们在此仍需要再次说明：由于数据样本所具有的代表性以及局限性，如果有朋友希望引用此书的某些数据时，特别建议将书中数据的上下文之间的关系进行说明，或者可以从原始的数据中进行相应分析。通过相互比对，一方面验证算法的有效性，另一方面也可以相互进行实验佐证，利用尽可能合理与科学的方式来面对现实

的数据。对于断章取义的人而言，数据的真实价值会受到实际情况的影响，甚至可能会变成传播谣言的利器，我们在此特别声明，本书的分析基于实际的网络数据，主要的目标是利用技术手段客观地反映社会现象，最终实现让社会也可以“智能化地计算”，这也是西安交通大学社会智能与复杂数据处理实验室的核心使命之一。

感谢的人与话实在太多，无论是国家基金项目还是身边众多的老师与朋友，每个鼓励与支持，每个批评与指责，都是我们努力与前进的动力。同时也要感谢这一群有朝气、有想法、勇于付出的青年才俊们，也正是他们长期的努力、默默地付出与辛勤地劳作，才换来了实验室一个又一个的喜报与阶段成绩，在他们的努力与快速成长下，也使我打破了“人过四十不学艺”旧例，虽然“已早生华发”仍需要不断地竭尽全力地努力，希望自己以及自己的学生们不仅可以成为一个“不器”的君子，也应该具有“至善”的品质。

在此，还需要再次感谢的是我的好友张澄宇博士专程为我们实验室书写的“君子不器”，以及青岛宝韵亦尚文化传媒公司的官爱兵董事长为我们题写的“止于至善”，这两幅字目前已悬挂在我们的实验室内，成为了我们实验室共同的精神资产与文化遗产。

借字以明志。

卡内基梅隆大学  
语言技术研究所（LTI）  
2015年7月10日